

University of Dundee

## Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals

23andMe Research Team; Lee, James J.; Wedow, Robbee; Okbay, Aysu; Kong, Edward; Maghzian, Omeed

*Published in:*  
Nature Genetics

*DOI:*  
[10.1038/s41588-018-0147-3](https://doi.org/10.1038/s41588-018-0147-3)

*Publication date:*  
2018

*Document Version*  
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

### *Citation for published version (APA):*

23andMe Research Team, Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., ... Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50(8), 1112-1121. <https://doi.org/10.1038/s41588-018-0147-3>

### **General rights**

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Supplementary Note for

Gene discovery and polygenic prediction from a 1.1-million-person GWAS  
of educational attainment

Correspondence to: [daniel.benjamin@gmail.com](mailto:daniel.benjamin@gmail.com), [a.okbay@vu.nl](mailto:a.okbay@vu.nl),  
[peter.visscher@uq.edu.au](mailto:peter.visscher@uq.edu.au)

### **This PDF file includes:**

Supplementary Text  
Supplementary Figures 1 to 29

## Table of Contents

<b>SUPPLEMENTARY TEXT</b>	<b>5</b>
<b>1. Methods GWA Studies</b>	<b>5</b>
1.1. Study Overview	5
1.2. Cohorts in <i>EduYears</i> Meta-Analysis	5
1.3. Phenotypes	6
1.4. Genotyping and Imputation	6
1.5. Association Analyses	6
1.6. Quality Control	7
1.7. Additional Diagnostics	7
1.8. <i>EduYears</i> Meta-Analysis ( $N = 1,131,881$ )	8
1.9. Clumping Algorithm and Definition of Lead SNPs	11
1.10. Replication of EA2 Lead SNPs	13
1.11. <i>Cognitive Performance, Math Ability and Highest Math</i>	15
1.12. Association Analyses of <i>CP, Math Ability and Highest Math</i>	16
1.13. MTAG of <i>CP, EduYears, Math Ability, and Highest Math</i>	17
1.14. Credibility of MTAG-Identified Lead SNPs	18
<b>2. Within-Family Association Analyses</b>	<b>21</b>
2.1. Introduction	21
2.2. Within-Family Association Analyses	21
2.3. Selection of SNPs Analyzed in Within-Family Tests	22
2.4. Calculating a Theoretical Benchmark for Within-Family Association Results	23
2.5. Winner's Curse Adjustment	23
2.6. Calculating Theoretical Benchmarks	24
2.7. Sign Tests	26
2.8. Within-Family Regression Test	28
2.9. Discussion and Additional Analyses	30
2.10 Appendix: Derivation of Adjustment for Assortative Mating	35
2.11 Appendix: Unified Regression Analyses	38
2.12 Appendix: LD Score Regression and Childhood Rearing Environment	38
<b>3. Heritability and Genetic Correlation Across Cohorts</b>	<b>41</b>
3.1 Introduction	41
3.2 How Population Differences Affect Cross-Cohort Prediction Accuracy	42
3.3 Variation in Heritability and Mean Genetic Correlation of <i>EduYears</i> Across Cohorts	43
3.4 Observed Cohort Characteristics	45
3.5 Cohort Characteristics and Heritability of <i>EduYears</i>	48
3.6 Cohort Characteristics and Genetic Correlation of <i>EduYears</i>	49
3.7 UK Biobank Analyses	51
3.8 Concluding Discussion	53
<b>4. X-Chromosome Analysis</b>	<b>57</b>
4.1 Introduction	57
4.1 Notation and Theoretical Framework	57
4.2 UK Biobank: Imputation, Quality Control and Association Analyses	61
4.3 UK Biobank Association Results	63

4.4 Association Analysis in 23andMe .....	64
4.5 Quality Control of UK Biobank and 23andMe Results .....	64
4.6 Meta-Analysis of UK Biobank and 23andMe Results ( $N = 694,894$ ) .....	64
4.7 Comparison to Autosomes.....	65
<b>5. Biological Mechanisms .....</b>	<b>69</b>
5.1. Introduction .....	69
5.2. Methods: Enriched Tissues/Cell Types, Enriched Gene Sets, Causal Genes, BrainSpan Developmental Transcriptome .....	69
5.3. Methods: Robustness Checks of Causal Genes and Enriched Gene Sets.....	73
5.4. Methods: Causal SNPs.....	77
5.5. Results: Enriched Tissues/Cell Types.....	80
5.6. Results: Causal Genes and Enriched Gene Sets .....	82
5.7. Results: Causal SNPs .....	119
5.8. Omnigenicity .....	121
<b>6. Prediction.....</b>	<b>124</b>
6.1 Introduction .....	124
6.2 Constructing Polygenic Scores .....	124
6.3 Defining Prediction Accuracy .....	126
6.4 GWAS- <i>EduYears</i> Polygenic Score .....	126
6.5 MTAG-Based Polygenic Scores.....	133
6.6 Comparing Observed Gains in Prediction Accuracy to Theoretical Predictions.....	135
6.7 Comparing Trait-Specific Scores.....	136
<b>7. Contributions and Acknowledgements .....</b>	<b>138</b>
7.1 Author Contributions .....	138
7.2 Cohort Contributions .....	139
7.3 Additional Acknowledgements .....	139
7.4 Extended Acknowledgements.....	146
<b>8. References .....</b>	<b>149</b>
<b>SUPPLEMENTARY FIGURES .....</b>	<b>168</b>
Supplementary Figure 1. Quantile-quantile Plots from Meta-analysis of <i>EduYears</i> ( $N = 1,131,881$ )....	169
Supplementary Figure 2. LD Score Plot from Meta-analysis of <i>EduYears</i> ( $N = 1,131,881$ ).....	170
Supplementary Figure 3. Replication of EA2 Lead SNPs.....	171
Supplementary Figure 4. Testing for Heterogeneous Effects of Lead SNPs. ....	172
Supplementary Figure 5. Meta-Analysis of X-Chromosomal SNPs ( $N = 694,894$ ). ....	173
Supplementary Figure 6. Comparison of Autosomal and X-Chromosomal Association Results. ....	174
Supplementary Figure 7. Flowchart of Biological Annotation. ....	175
Supplementary Figure 8. Roles of Selected Newly Prioritized Genes in Neuronal Communication. ....	176
Supplementary Figure 9. Regional Association plots for Four Likely Causal SNPs Identified using CAVIARBF. ....	177
Supplementary Figure 10. Predictive Power of Polygenic Score as a Function of Pruning at Different $P$ Value Thresholds. ....	178
Supplementary Figure 11. Mean Prevalence of Schooling Outcomes by <i>EduYears</i> PGS Quintile. ....	179
Supplementary Figure 12. Predictive Power of GWAS- <i>EduYears</i> Polygenic Score Compared to Other Variables (top) and as Attenuated by Additional Controls (bottom).....	180
Supplementary Figure 13. Polygenic Score Prediction in <i>Add Health</i> and <i>HRS</i> . ....	181
Supplementary Figure 14. Manhattan Plot for <i>Cognitive Performance</i> ( $N = 257,841$ ).....	182
Supplementary Figure 15. Manhattan Plot for <i>Self-Rated Math Ability</i> ( $N = 564,698$ ).....	183

Supplementary Figure 16. Manhattan Plot for <i>Highest Math</i> ( $N = 430,445$ ). .....	184
Supplementary Figure 17. Inverted Manhattan Plot of GWAS and MTAG results for <i>EduYears</i> . .....	185
Supplementary Figure 18. Inverted Manhattan Plot of GWAS and MTAG Results for <i>Cognitive Performance</i> . .....	186
Supplementary Figure 19. Inverted Manhattan Plot of GWAS and MTAG Results for <i>Math Ability</i> . ....	187
Supplementary Figure 20. Inverted Manhattan Plot of GWAS and MTAG Results for <i>Highest Math</i> . ...	188
Supplementary Figure 21. Summary Overview of $m_c$ Estimates in Sibling Cohorts. ....	189
Supplementary Figure 22. Brain-Specific Expression of Significantly Enriched Gene Sets across Development. ....	190
Supplementary Figure 23. DNase I Hypersensitivity in Fetal Tissues/Cell Types as a Predictor of SNP Effects on <i>EduYears</i> . ....	191
Supplementary Figure 24. Heritability Enrichment of Genes That Are Broadly or Specifically Expressed. ....	192
Supplementary Figure 25. Binary Gene Sets with Strongest and Weakest Heritability Enrichment (15 of Each). ....	193
Supplementary Figure 26. Predictive Power of Polygenic Score as a Function of the Size of the <i>EduYears</i> GWAS Discovery Sample. ....	194
Supplementary Figure 27. Predictive Power of Chromosome-Specific <i>EduYears</i> Polygenic Scores in <i>Add Health</i> and <i>HRS</i> . ....	195
Supplementary Figure 28. Predictive Power of Chromosome-Specific <i>EduYears</i> Polygenic Scores in Sample-Size Weighted Meta-Analysis of <i>Add Health</i> and <i>HRS</i> . ....	196
Supplementary Figure 29. Comparison MTAG PGSSs Based on Trait-Specific MTAG Association Statistics and MTAG Association Statistics for Other Traits. ....	197

## Supplementary Text

### 1. Methods GWA Studies

#### **1.1. Study Overview**

Our primary analysis extends the discovery sample of a previous genome-wide association study (GWAS) of educational attainment<sup>1</sup> from  $N = 405,072$  to  $N = 1,131,881$  individuals. We also conducted genome-wide association analyses of cognitive performance ( $N = 257,841$ ), self-reported *Math Ability* ( $N = 430,445$ ) and *Highest Math* class ever successfully completed ( $N = 564,698$ ). In what follows, we refer to the four variables as *EduYears*, *CP*, *Math Ability* and *Highest Math*.

Below, we begin by describing the methods used in our primary GWAS of *EduYears* and summarize its key findings. Next, we describe the GWASs of *CP*, *Math Ability* and *Highest Math*, all of which were performed using protocols designed to be as similar as possible to that of the primary GWAS. We conclude the section by describing a joint analysis of the four traits that exploits their substantial genetic correlations to further improve both the predictive power of polygenic scores based on our results and our power to detect individual genetic associations.

#### **1.2. Cohorts in *EduYears* Meta-Analysis**

In this study, we meta-analyzed summary statistics from 71 separate genome-wide association studies of educational attainment. Our analyses extend a previous genome-wide study of educational attainment<sup>1</sup> (referred to as EA2 in what follows), which combined data from 64 discovery cohorts and one replication cohort, yielding a combined sample size of  $N = 405,072$ . The EA2 study, in turn, built on an earlier GWAS (which we call EA1)<sup>2</sup>.

Relative to EA2, we augmented the sample size in two ways. First, we replaced some EA2 cohort-level results files with results from the cohort based on new analyses of larger samples. Doing so was possible for some EA2 cohorts for which expanded genotyped samples became available after the discovery stage of EA2 was closed. Second, we added data from new cohorts that did not contribute to EA2. **Supplementary Table 16** provides summary information about the 12 cohorts that contributed new data for the present study (for analogous information about the EA2 cohorts, see Supplementary Table 16 of Okbay et al.<sup>1</sup>). Our final meta-analysis also includes 59 of the 65 original EA2 cohorts (the table caption of **Supplementary Table 16** lists the six EA2 cohorts whose results were replaced with results from a larger sample).

By meta-analyzing summary statistics from association analyses conducted in the 59 EA2 cohorts (combined  $N = 199,819$ ), and the twelve cohorts in **Supplementary Table 16** (combined  $N = 932,062$ ), we obtain our final discovery sample of  $N = 1,131,881$ . Over half of the increase in sample size relative to EA is due to sample-size increases in the 23andMe cohort (an increase from  $N = 76,155$  in EA2 to  $N = 365,536$ ) and UKB (increase from 111,349 to 442,183).

The lead PI of each cohort affirmed that the results contributed to the study were based on analyses approved by the local Research Ethics Committee and/or Institutional Review Board responsible for overseeing research.

### 1.3. Phenotypes

The study-specific phenotype measurements and distributions for the new cohorts are summarized in **Supplementary Table 17** (for analogous information about the EA2 cohorts, see Supplementary Table 18 in Okbay et al.<sup>1</sup>). As in our prior work<sup>1,2</sup>, we map each major educational qualification that can be identified from the cohort’s survey measure to an International Standard Classification of Education (ISCED) category. To construct our outcome variable, *EduYears*, we impute a years-of-education equivalent for each ISCED category. Across all cohorts, the sample-size-weighted mean of *EduYears* is 16.8 years of schooling with a standard deviation of 4.2.

### 1.4. Genotyping and Imputation

**Supplementary Table 18** reports information about genotyping platform, pre-imputation quality-control filters applied to the genotype data, subject-level exclusion criteria, imputation software used, and the reference sample used for imputation in each of the new cohorts. Imputation was conducted using a reference panel from either the 1000 Genomes Project<sup>3</sup> or a larger panel subsequently released by Haplotype Reference Consortium<sup>4</sup>.

### 1.5. Association Analyses

Cohorts were asked to estimate this regression equation for each measured SNP:

$$EduYears = \beta_0 + \beta_1 SNP + \mathbf{PC} \boldsymbol{\gamma} + \mathbf{B} \boldsymbol{\alpha} + \mathbf{X} + \epsilon, \quad (1.1)$$

where *SNP* is the allele dose of the SNP; **PC** is a vector of the first ten principal components of the variance-covariance matrix of the genotypic data, estimated after the removal of genetic outliers (we instead used twenty principal components in UKB analyses); **B** is a vector of standardized controls, including a third-order polynomial in

year of birth, an indicator for being female, and their interactions; and  $\mathbf{X}$  is a vector of study-specific controls. Cohort analysts were asked to impose a number of standard subject-level filters prior to running the analyses. These include: (i) each subject's *EduYears* was measured at an age of at least 30, (ii) each subject passed the cohort's quality control, which always include the removal of genetic outliers and individuals with poor genotyping rates, and (iii) each subject is of European ancestry.

**Supplementary Table 19** provides study-specific details about the association analyses conducted in the new cohorts. Column 2 shows the association software used by each study analyst. Column 3 reports whether the cohorts omitted any of the basic control variables recommended in the Analysis Plan in their specification. Column 4 lists extra controls included by the cohorts in the vector  $\mathbf{X}$ , such as controls for cohort-specific events that may have impacted the education system in the cohort. Column 5 reports whether association analyses were conducted using mixed linear models that may yield more robust inference, especially in family-based samples. In the 23andMe sample, the association analyses were conducted in a sample of European-ancestry research participants selected so that in the sample, no pair of research participants share more than 700 cM identically by descent.

## 1.6. Quality Control

We applied the quality-control protocol and filters described in EA2<sup>1</sup> to the new results files. Several of the quality-control and filtering steps are implemented by the software EasyQC, using the 1000 Genomes Project<sup>3</sup> phase 1 European sample reference files provided on the EasyQC website.<sup>a</sup>

The main filtering steps involved dropping SNPs that: (i) are known to have strand issues in some imputation programs, (ii) have missing or incorrect numerical values supplied for some variables (e.g., a  $P$  value of association outside the range 0 to 1), (iii) have a minor allele count below 25, (iv) have poor imputation accuracy, (v) are indels or not located on the autosomes, or (vi) have invalid or duplicated chromosomal coordinates or whose alleles do not match those in the reference file. In association results from analyses of the full release of the UK Biobank data, we further filter out all SNPs that are not in the Haplotype Reference Consortium's reference panel.

## 1.7. Additional Diagnostics

---

<sup>a</sup><http://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/>



After applying the filters described in the previous section, we conducted several additional diagnostic checks before clearing a cohort-level results file for inclusion in the meta-analysis.

The first four of these diagnostics are graphical and summarized below.

*Allele Frequency Plots (AF Plots):* We looked for errors in allele frequencies and strand orientations by visually inspecting a plot of the sample allele frequency of filtered SNPs against the frequency in the 1000 Genomes phase 1 version 3 European panel<sup>3</sup>.

*P value vs Z statistic Plots (PZ Plots):* We verified that the reported *P* values are consistent with the *P* values implied by the coefficient estimates and standard errors in the results file.

*Quantile-Quantile Plots (QQ Plots):* We visually inspected the cohort-level QQ plots to look for evidence of unaccounted-for stratification.

*Predicted vs Reported Standard-Error Plots (PRS Plots):* We investigated if the standard errors reported in the files are approximately consistent with the reported sample size, allele frequency, and phenotype distribution. For a random subset of 500,000 SNPs, we also plotted the predicted standard errors against the actual standard errors reported by the cohort.

We generated the above four plots for each new results file and inspected them for anomalies. Potential issues were discussed with cohort-level analysts and sometimes resulted in re-uploading of results. We also used bivariate LD Score regression to verify that the estimated genetic correlations between all large cohorts (defined as  $N > 10,000$ ) were large and positive.

All of our final analyses are based on results files that pass all the diagnostic tests described above.

### **1.8. EduYears Meta-Analysis ( $N = 1,131,881$ )**

We use the software program METAL<sup>5</sup> to conduct sample-size-weighted meta-analysis of all SNPs that passed the quality-control thresholds in the 71 results files. Applying a sample-size filter of 500,000 leaves us with meta-analysis results for 10.02M autosomal SNPs.

**Supplementary Figure 1** reports quantile-quantile plots of the *P* values from the meta-analysis. **Panel a** shows the overall distribution of *P* values, and **Panel b** shows the *P* values of SNPs categorized by allele frequency. As expected under polygenicity<sup>6</sup>, the *P* values deviate strongly from a uniform distribution ( $\lambda_{GC} = 2.04$ ). The strength of this deviation depends strongly on allele frequency. For common variants (defined as having

a minor allele frequency above 5%) the genomic control factor is  $\lambda_{GC} = 2.85$ ; for low-frequency variants (MAF 1-5%) we have  $\lambda_{GC} = 1.63$ ; and finally, for rare variants (MAF < 1%) we have  $\lambda_{GC} = 1.20$ .

We did not apply cohort-level genomic control<sup>7</sup> to the cohort-level results files prior to meta-analysis. Instead, we meta-analyzed unadjusted cohort-level summary statistics and subsequently inflated the standard errors from the meta-analysis by the square root of the intercept ( $\sqrt{1.11}$ ) from an LD Score regression<sup>8</sup>. As recommended by the developers of LD Score regression<sup>8</sup>, the LD Scores are estimated using the 1000G reference sample, even though LD Score regression is run restricting the sample to just the HapMap3 SNPs with allele frequency above 1%. (The developers recommend restricting to these HapMap3 SNPs because they can be well imputed.) Consequently, the intercept is estimated from the HapMap3 SNPs rather than the full set of 1000G SNPs contained in the GWAS results. When we use this intercept to inflate the standard errors in the GWAS results, we are implicitly assuming that the inflation of the  $z$ -statistics due to stratification or other biases in the HapMap3 SNPs is the same as for the 1000G SNPs. The primary difference between HapMap3 SNPs and 1000G SNPs is that there are many more SNPs with rare alleles in the set of 1000G SNPs. Thus, the LD Score intercept adjustment could lead to an inflated Type-I error rate for rare SNPs if the bias for rare SNPs were greater than that of common SNPs. We are not aware of any evidence on this point. We note, however, that of the 1,271 lead SNPs identified for *EduYears*, only two have an allele frequency less than 1%: rs186456786 and rs182355396, which have minor allele frequencies of 0.94% and 0.84%, respectively. Since there are only two such SNPs, since both have minor allele frequencies close to 1%, and since the regression includes HapMap3 SNPs with minor allele frequencies as low as 1%, we think that any potential inflation of the Type-I error rate in our results due to the restriction of SNPs included in the LD Score regression is likely to be negligible.

Our LD Score intercept estimate of 1.11 is in the upper range of estimates previously reported in the literature (e.g., range 0.980 to 1.149 reported by Bulik-Sullivan et al.<sup>8</sup> for 24 traits). However, to be informative about the overall amount of inflation due to cryptic relatedness or stratification biases, the intercept must be interpreted relative to the overall observed inflation, which greatly exceeds that of the previous studies analyzed by Bulik-Sullivan et al.<sup>8</sup>. In our data, the average  $\chi^2$  test statistic among HapMap3 SNPs is 3.81, substantially larger than the range of values (1.033-1.802) for the set of traits analyzed by Bulik-Sullivan et al.<sup>8</sup>. Thus, the intercept estimate of 1.11 suggests that biases explain only a small share of the overall inflation in the test statistics, with roughly 95% of the inflation due instead to polygenicity. For a graphical summary of the LD Score stratification analyses, see **Supplementary Figure 2**. In **Supplementary Section 2**, we

report results from within-family association analyses that provide complementary evidence about the amount of stratification bias in our coefficient estimates. **Figure 1** displays the Manhattan plot for the main meta-analysis.

**Supplementary Table 2** shows the association results for the 1,271 approximately independent SNPs that reached genome-wide significance in our *EduYears* meta-analysis (see **Supplementary Section 1.3** for details on the clumping algorithm). In what follows, we refer to these approximately independent SNPs as our lead SNPs. Of the 1,271 lead SNPs, 1,190 have minor allele frequencies above 5%, and all but two of the remaining 81 SNPs have a minor allele frequency below 5%. A test of homogenous effects across cohorts fails to reject the null at the Bonferroni-adjusted  $P$  value threshold of  $0.05/1,271$  for all SNPs barring one. Nevertheless, **Supplementary Figure 4** shows that there is a tendency for the overall  $P$  value distribution from the tests of homogenous effects to deviate from the theoretical expectation. The 1,271 lead SNPs are spread across the autosomes, with a randomly selected SNP having a 17% chance of being in a genome-wide significant locus (i.e., pairwise  $r^2 > 0.1$  with at least one lead SNP).

We generated an omnibus test statistic for heterogeneity by summing the Cochran Q-statistics for heterogeneity across all 1,271 lead SNPs<sup>9</sup>. Because the software used for meta-analysis does not report Q-statistics, we inferred these values based on the reported heterogeneity  $P$  values. To do so, we treated each lead SNP as if it were available for each of the 71 cohorts in the meta-analysis, which implies that the Q-statistic for each lead SNP has a  $\chi^2$  distribution with 70 degrees of freedom. The resulting Q-statistics were almost perfectly consistent with the  $l^2$  value reported by METAL, suggesting that our approximation is appropriate. The sum of these Q-statistics is therefore (approximately)  $\chi^2$ -distributed with  $70 \times 1,271 = 88,970$  degrees of freedom. This gave us an omnibus Q-statistic of 91,830, with corresponding  $P$  value equal to  $9.68 \times 10^{-12}$ . We return to the issue of heterogeneous effects across cohorts in **Supplementary Section 3**.

To gauge the magnitude of the estimated SNP effects, we used a well-known approximation to transform the  $z$ -statistics from the sample-size-weighted meta-analysis (the output of the software METAL) into unstandardized regression coefficients:

$$\hat{\beta}_j \approx z_j \frac{\hat{\sigma}_Y}{\sqrt{2N_j \text{MAF}_j (1 - \text{MAF}_j)}} \quad (1.2)$$

for SNP  $j$  with minor allele frequency  $\text{MAF}_j$ , sample size  $N_j$ ,  $z$ -statistic  $z_j$ , and standard deviation of the phenotype  $\hat{\sigma}_Y$ . For a derivation, see the SOM in EA1<sup>2</sup>. Without adjustment for winner's curse, the estimated effects (in absolute value) of the 1,271

genome-wide significant SNPs are all in the range 0.008-0.053 SD units, corresponding to approximately 2 to 9 weeks of schooling per reference allele (assuming the standard deviation of *EduYears* is 40.2). An additional copy of the trait-increasing allele is associated with 2.7 weeks of schooling on average. When we consider common and low-frequency variants separately, the averages are instead 2.5 and 6 weeks.

We also used an empirical Bayes' framework to calculate winner's-curse-adjusted estimates of the effect sizes. The framework in its entirety is described in Section 2.5, but its key assumption is that SNP effects are drawn randomly from the following mixture distribution:

$$\beta_j \sim \begin{cases} N(0, \tau^2) & \text{with probability } \pi \\ 0 & \text{otherwise,} \end{cases} \quad (1.3)$$

We used the summary statistics from the full meta-analysis to obtain estimates of the fraction of non-null variants ( $\pi$ ) and the variance of non-null SNP effect sizes ( $\tau^2$ ). The resulting estimates are  $(\hat{\tau}^2, \hat{\pi}) = (3.98 \times 10^{-6}, 0.60)$ , from which we calculated posterior distributions for the effect sizes of all lead SNPs. We found that after adjusting for winner's curse, the average effect falls from 2.7 weeks per allele to 1.8 weeks, and no longer varies appreciably between low-frequency (1.6 weeks) and common variants (1.8 weeks). We also used the posterior distributions to calculate the total fraction of variation accounted for by the lead SNPs. The predicted fraction of variance explained by the variants jointly, 3.86%, is broadly consistent with our estimates of the predictive power of polygenic scores based on lead SNPs in independent validation samples.

## 1.9. Clumping Algorithm and Definition of Lead SNPs

Here, we assess the sensitivity of our conclusions about the number of lead SNPs to alternative definitions and to the choice of the reference file used to estimate LD structure.

To examine the sensitivity of our conclusions to the choice of reference file, we conducted follow-up analyses in a sample of approximately unrelated individuals (pairwise relatedness  $< 0.025$ ) of European ancestry ( $N = 405,519$ ) from *UKB*. Our first analysis is motivated by the fact that the 1,271 lead SNPs were identified using noisily estimated patterns of LD from a small reference file ( $N = 294$ ). Estimation error could inflate the number of SNPs classified as approximately independent. We therefore reran our clumping algorithm using the *UKB* reference sample, which is large enough to ensure that estimation error will be negligible. With this alternative reference sample, we identified 1,223 approximately independent SNPs at genome-wide significance, suggesting that our conclusions about the number of lead SNPs are not sensitive to the

exact reference sample of European-ancestry individuals used to estimate linkage disequilibrium.

There is no universally agreed-upon clumping algorithm, and different studies often use slightly different ones. For example, some algorithms do not assign *all* SNPs whose pairwise  $r^2$  with the lead SNP exceeds the user-specified cutoff to the lead SNP's clump, but only the subset of these SNPs whose distance to the lead SNP is below some cutoff (e.g. 250 kb). As another example, some algorithms involve a second stage in which lead SNPs that are physically close to each other are sometimes merged and considered to be a single locus. Since our clumping procedure did not involve a physical distance threshold, such merging is likely to be very conservative in our setting. If we apply the clumping algorithm used in a recent study of schizophrenia<sup>10</sup>, our lead SNPs span 795 loci.

We also supplemented our main analyses with a conditional and joint multiple-SNP analysis (COJO)<sup>11</sup> using summary statistics from the main meta-analysis. In our COJO analysis, we again use the *UKB* reference sample, primarily to ensure that LD structure is estimated with minimal error. But since *UKB* comprises a large share of our overall estimation sample, a potential additional advantage is that its LD structure may be more representative of the discovery sample, which would improve the accuracy of the COJO estimator.

Before running COJO, we applied recommended SNP filters<sup>11</sup>, excluding SNPs with (i)  $MAF < 0.01$ , (ii) imputation  $r^2$  of  $< 0.3$ , (iii) HWE  $P$  value of  $< 10^{-6}$ , or (iv) missingness rate of  $> 5\%$ . These restrictions leave us with  $\sim 4.9$ M SNPs. We performed COJO using the implementation found in the GCTA software (Version 1.90.0 beta). Model selection was performed using the stepwise selection process outlined in the original COJO paper<sup>11</sup> in which SNPs from across the genome are iteratively added to the model. We set the LD window to 100 Mb, i.e., SNPs that are further than 100 Mb are assumed to have zero LD correlation.

Our COJO analysis identified 765 variants at genome-wide significance. In evaluating how this estimate compares to the number of lead SNPs identified by our clumping algorithm, it is important to consider that the COJO analysis was based on a restricted set of SNPs. To aid comparisons, we applied our clumping algorithm to the  $\sim 4.9$ M SNPs that passed COJO filters and found 1,053 lead SNPs when *UKB* is used as the reference sample (1,070 if we instead use the smaller reference sample used in our primary analyses).

A priori, it was not clear whether the number of COJO-identified associations should have been expected to be larger or smaller than the number of approximately independent SNPs identified by our clumping algorithm (for a fixed set of SNPs). On the

one hand, COJO could increase the number if there are many loci harboring secondary associations. On the other hand, COJO could reduce the number of associations if there are many pairs of lead SNPs with pairwise  $r^2$  just below 0.1 and  $P$  values just below the genome-wide significance cutoff. If so, conditioning on one of the SNP may increase the association  $P$  value of the other so that it ceases to be genome-wide significant. Empirically, our findings suggest that the second effect appears to dominate.

We classified each of the 765 COJO hits as either primary or secondary by applying our clumping algorithm to the list of COJO variants (**Supplementary Table 3**), again using *UKB* as our reference sample and an  $r^2$  threshold of 0.1. We found that our clumping algorithm eliminated 60 SNPs from the original list of 765 COJO hits (pairwise  $r^2 > 0.1$  with at least one COJO variant). We call these 60 variants secondary associations and the remaining 705 variants primary associations.

### 1.10. Replication of EA2 Lead SNPs

Okbay et al. reported a replication analysis of the 74 lead SNPs from their discovery meta-analysis in an independent replication sample. Here we conduct a replication analysis of the 162 lead SNPs identified at genome-wide significance in Okbay et al.’s pooled (discovery and replication) meta-analysis ( $N = 405,073$ ).

Of the 162 SNPs, 158 of them pass quality-control filters in our updated meta-analysis, so we focus on those. To examine their out-of-sample replicability, we calculated Z-statistics from the subsample of our data ( $N = 726,808$ ) that was not included in EA2. Let the Z-statistics of association from, respectively, EA2, the new data, and our final EA3 meta-analysis, be denoted by  $Z_1$ ,  $Z_2$  and  $Z$ . Note that we cannot calculate  $Z_2$  directly. For example, 23andMe contributed association results for  $N = 76,155$  individuals to EA2 and results for  $N = 365,538$  individuals to EA3. The first cohort is a subset of the second, but we do not have summary statistics from association analyses conducted only in subjects that contributed to the second but not the first meta-analysis. The same is true for the other cohorts listed in Supplementary Table 16 that increased their sample sizes from EA2 to EA3, with the exception of the UKB (for which we have access to individual-level data).

However, we can calculate  $Z_2$  indirectly. Since our meta-analysis used sample-size weighting<sup>5</sup>,  $Z_2$  is implicitly defined by:

$$Z = \sqrt{\frac{N_1}{N}} Z_1 + \sqrt{\frac{N_2}{N}} Z_2,$$

where SNP subscripts have been dropped for notational convenience and  $N$ 's are sample sizes. Because this formula holds when  $Z_1$  and  $Z_2$  are independent, the implicitly-defined  $Z_2$  is interpreted as the additional information contained the new data. If in fact the EA2 and the new data were correlated (because, for example, the additional 23andMe data likely contains individuals related to individuals in the earlier 23andMe data),  $Z_2$  would not be equal to the  $Z$ -statistic of association calculated directly in the new data. However, the implicitly-defined  $Z_2$  is the correct value to use for assessing replicability of the EA2 results because it captures the independent component of the new data.<sup>b</sup>

Of the 158 SNPs, we find that 154 have matching signs in the new data (for the remaining four SNPs, the estimated effect is never statistically significant at  $P < 0.10$ ). Of the 154 SNPs with matching signs, 143 are significant at  $P < 0.01$ , 119 are significant at  $P < 10^{-5}$ , and 97 are significant at  $P < 5 \times 10^{-8}$ . The replication results are shown graphically in **Supplementary Figure 3**.

To help interpret these results, we used the statistical framework from Section 1.8 in the Supplementary Methods of Okbay et al.<sup>1</sup> to calculate the expected replication record under the null that all 158 SNPs are true associations. Below, we show the expected replication record (mean and standard deviation of the number of successful replications in each category). The theoretical projections are based on shrinkage parameters estimated from EA2 summary statistics (used to adjust the EA2 effect sizes for winner's curse):  $(\hat{\tau}^2, \hat{\pi}) = (5.02 \times 10^{-6}, 0.33)$ . To facilitate comparisons, the empirical findings mentioned above that correspond to each prediction are provided in this table:

Test	Theoretical Expected Replication		Observed Replication
	Mean	Standard deviation	
Sign Concordance	157.7	0.04	154
$P < 0.01$	152.3	0.62	143
$P < 10^{-5}$	118.3	0.04	119
$P < 5 \times 10^{-8}$	79.4	3	97

<sup>b</sup> We note, however, that in our application, the implicitly-defined  $Z_2$  is approximate because the equation is only exact if identical methods are used at the cohort level to adjust for stratification biases. In reality, cohort-level genomic control was applied in EA2, whereas the current study used LD Score adjustment. However, Okbay et al. showed in a robustness analysis that in their setting, differences between the two approaches are negligible (see footnote g in the Supplementary Methods of Okbay et al.<sup>1</sup>).

For the first two categories, the empirical replication record is weaker than predicted; for the third category, the empirical replication record is close to the theoretical prediction; and for the fourth category, the empirical replication record is stronger than predicted. However, the quantitative discrepancies are always fairly small. The weaker-than-expected success rates observed in the first two categories is likely due to a mix of two factors: (i) imperfect genetic correlation between the cohorts used to estimate  $Z_1$  and  $Z_2$ , and (ii) a low, but non-zero, false-discovery rate in EA2 (our theoretical calculations assume all 158 SNPs are non-null).

### 1.11. Cognitive Performance, Math Ability and Highest Math

**Supplementary Table 40** provides summary information about the phenotypes used in our remaining three genome-wide association analyses, two of which were conducted exclusively among research participants of the personal genomics company 23andMe who answered survey questions about their mathematical background. Our first variable, *Math Ability* ( $N = 564,698$ ), is derived from the respondent’s answer to the categorical question “How would you rate your mathematical ability?” Our second variable, *Highest Math* ( $N = 430,445$ ), is similarly derived from the answer to a question about the most advanced math course ever successfully completed. **Supplementary Table 40** provides additional information about the questions used and how we generated a quantitative variable by mapping each response category to a numerical value.

In our analyses of *CP*, we partnered with investigators from a published study of general cognitive ability in European-ancestry individuals ( $N = 35,298$ ) conducted by the COGENT consortium<sup>12</sup>. We combined summary statistics from their study with new genome-wide association analyses of cognitive performance in the UK Biobank ( $N = 222,543$ ). Our final analyses are based on a sample-size weighted meta-analysis of these two results files ( $N = 257,841$ ).

In the COGENT study all 35 participating cohorts analyzed a phenotype defined as the first principal component derived from three or more neuropsychological tests (the exact tests varied by cohort as described in the original study). In the *UKB* analyses, following prior work, we used the respondent’s score on a test of verbal-numerical reasoning as our cognitive phenotype.<sup>c</sup> The test, which contains thirteen logic and

---

<sup>c</sup> Other work has found that the genetic correlation between this measure and general intelligence among children is approximately 0.8<sup>327</sup>. The *UKB* contains data on three other cognitive tests, in addition to the one we use: reaction time, pairs matching, and prospective memory. A previous study contains a detailed description of all four cognitive tests in the *UKB* and a comprehensive analyses of their psychometric properties



reasoning questions with a two-minute time limit, was designed as a measure of fluid intelligence. Each respondent took the test up to four times. Our phenotype is the mean of the standardized score across the occasions on which the respondent took the test.

### 1.12. Association Analyses of *CP*, *Math Ability* and *Highest Math*

Our new genome-wide analyses of cognitive performance in *UKB* were conducted using methods identical to those described above for *EduYears* in *UKB*, albeit in a smaller sample because the phenotype is only available for approximately one half of the respondents. Our genome-wide association analyses of *Math Ability* and *Highest Math* in 23andMe were also conducted using methods identical to those described for *EduYears* in 23andMe. Thus, for details on genotyping and imputation and association models used in our additional genome-wide analyses in *UKB* and 23andMe, we refer to **Supplementary Tables 17** and **18**. Before conducting clumping and further analyses, we applied the same set of quality-control filters to all results files and verified that no file failed any of the diagnostic checks described in Section 1.7.

We performed a sample-size-weighted meta-analysis of the *UKB* and COGENT summary statistics, imposing a minimum-sample-size filter of 100,000, leaving 10.10M SNPs in our final results file. For *Math Ability* and *Highest Math*, we imposed sample-size filters of 500,000 and 350,000, respectively. Applying these final filters leaves us with association statistics for ~10.0M SNPs in each of the 23andMe results files.

To reduce stratification biases, we adjusted the test statistics using the estimated intercepts from LD Score regressions. The three estimated intercepts are all small relative to the overall inflation in test statistics: 1.073 for *Math Ability*, 1.105 for *Highest Math*, and 1.046 for *CP*. Applying our clumping algorithm to the *P* values derived from the adjusted test statistics, we identified 618 approximately independent SNPs associated with *Math Ability* (**Supplementary Table 11**), 365 associated with *Highest Math* (**Supplementary Table 12**), and 225 SNPs associated with *CP* (**Supplementary Table 13**). Manhattan plots from the analyses are shown in **Supplementary Figures 14** (*CP*), **15** (*Math Ability*) and **16** (*Highest Math*).

Although we have considered each of these phenotypes separately and therefore used the standard genome-wide significance *P* value threshold of  $5 \times 10^{-8}$  to identify the lead SNPs for each phenotype, it may also be of interest to consider them jointly with a Bonferroni corrected *P* value threshold of  $1.25 \times 10^{-8}$ . Using that threshold in the clumping

---

<sup>355</sup>. The study finds that the correlations between the four tests are relatively low, and the verbal-numerical reasoning test has the highest retest reliability, estimated to be 0.65.

algorithm yields 1,024, 171, 492, and 284 lead SNPs for *EduYears*, cognitive performance, *Math Ability*, and *Highest Math*, respectively.

### 1.13. MTAG of *CP*, *EduYears*, *Math Ability*, and *Highest Math*

Next, we used a recently developed method<sup>13</sup> to perform a joint analysis of the four phenotypes. The method, MTAG, only requires summary from a GWAS of each trait and accounts for (possibly unknown) sample overlap through the use of bivariate LD Score regression<sup>14</sup>. Prior to our MTAG analysis, we applied the MTAG-recommended SNP filters, dropping SNPs with minor allele frequency below 1% and SNPs whose effects fail a sample-size cutoff (66.7% of the 90<sup>th</sup> percentile of the sample-size distribution). In total, ~7.1M SNPs satisfy these restrictions in all four results files, and the MTAG analyses that follow are all based on this restricted set of SNPs.

To motivate our application, **Supplementary Table 14** reports the estimated genetic correlations between our four phenotypes. The estimates are all substantial, ranging from 0.51 (*Math Ability* versus *EduYears*) to 0.85 (*Highest Math* versus *Math Ability*). The substantial genetic correlations suggest that a joint analysis of the phenotypes may have better power than single-phenotype analyses based on smaller samples.

**Supplementary Table 41** instead compares, for each of the four phenotypes, single-phenotype GWAS results along several dimensions. The comparisons are all based on the set of SNPs that passed MTAG filters. Overall, the results reported in the table conform to theoretical expectations, with the largest gains observed for *Highest Math* and *CP*. Following a methodology described in the original MTAG paper, we used the observed increases in average  $\chi^2$  to calculate how much larger the GWAS sample size would have to be to attain an equivalent increase in expected  $\chi^2$ . We find that the MTAG analysis of *EduYears*, *CP*, *Math Ability* and *Highest Math* yielded gains equivalent to augmenting the original sample sizes by 16%, 56%, 19% and 89%, respectively. Intuitively, our finding of the substantially larger gains for *Highest Math* than *Math Ability* is driven in large part by the higher genetic correlation between *Highest Math* and the *EduYears* and *CP* phenotypes (**Supplementary Table 14**). The number of new lead SNPs identified by MTAG is also greatest for the two phenotypes – *CP* and *Highest Math* – for which the calculations imply the largest increase in effective sample size. Information about all SNPs that reached genome-wide significance in the MTAG analysis is provided in **Supplementary Table 15**.

To allow graphical comparisons, **Supplementary Figures 17-20** show inverted Manhattan plots, with MTAG *P* values plotted above the x-axis, and the original GWAS

$P$  values below the x-axis. To facilitate comparisons, these comparative figures are all based on the 7.09M SNPs that passed MTAG filters.

#### 1.14. Credibility of MTAG-Identified Lead SNPs

The derivation of MTAG makes the important assumption that all SNP effects are drawn from a single variance-covariance matrix of effect sizes ( $\Omega$ ). Violations of this homogeneous- $\Omega$  assumption can inflate MTAG's false discovery rate (FDR). The problem arises when there are some SNPs that are null for a first trait but not a second trait, especially when the GWAS of the first trait is well-powered relative to the GWAS of the second trait. **Supplementary Table 44** shows that in our application, the *CP* GWAS had the lowest relative power, with a mean  $\chi^2$ -statistic 15-70% smaller than the remaining three phenotypes. Given the near tripling of the number of genome-wide significant lead SNPs, from 225 to 661, it is important to probe the credibility of the MTAG-identified lead SNPs. The most transparent way to do this is through replication in independent samples. As a robustness analysis, we therefore repeated our MTAG analysis, but this time omitting COGENT cohorts from the *CP* and *EduYears* GWASs (there is no known overlap between COGENT and the 23andMe GWASs and bivariate LD Score analysis failed to reject the null hypothesis of zero overlap). We again restricted the MTAG analysis to SNPs that pass recommended MTAG filters. Additionally, we limited the analysis to SNPs for which the COGENT results file contains summary statistics based on analyses of at least 25,000 individuals.

As expected, given that COGENT cohorts comprise a modest fraction of our total discovery samples, the restricted MTAG analysis described above yielded very similar findings. Applying our clumping algorithm to the summary statistics for *CP* by MTAG with COGENT cohorts omitted, we identified 578 approximately independent SNPs at genome-wide significance (compared to 661 in the original analysis). By construction, association statistics for all these SNPs are available in the COGENT results file (with  $N > 25,000$ ). As shown in **Supplementary Table 44**, 485 of the 578 SNPs (84%) had concordant signs in the independent COGENT sample, with 127 out of 578 significantly associated with *CP* at the 5% level (124 of them with the expected sign).

To evaluate this replication record, we used the statistical model described in Section 2.5 to calculate a posterior distribution for each genome-wide significant SNP's effect and used these to calculate the expected replication record in the COGENT data. Our calculations assume that the genetic correlation is perfect but take into account the fact that the estimated SNP heritability is approximately 60% higher in *UKB* than COGENT ( $h^2_{UKB}/SE_{UKB} = 0.23/0.01$ ;  $h^2_{COGENT}/SE_{COGENT} = 0.14/0.01$ ). Using

summary statistics from the *UKB* GWAS of *CP*, we estimated the fraction of non-null variants ( $\pi$ ) and the variance of non-null SNP effect sizes ( $\tau^2$ ).

From our estimates,  $(\hat{\tau}^2, \hat{\pi}) = (9.2 \times 10^{-6}, 0.46)$ , we calculated the posterior distribution of each SNP under the maintained assumption that it is non-null. We then estimated the expected replication record in the COGENT data by drawing simulated effects from the posterior distributions for each of the 578 SNPs, generating simulated COGENT summary statistics by adding simulated estimation error to the effect sizes, and calculating the observed replication record in the simulated summary statistics. This was repeated 1000 times, and we used the mean replication record of the simulated data as the expected replication record under the assumption that the MTAG results are true. Using this procedure, we found that in the COGENT replication sample, the expected number of correctly signed and nominally significant SNPs is 140.1 (SD = 10.1) and the expected number of SNPs with matching signs is 87.8% (SD = 1.3%). Overall, these analyses thus provide little evidence that MTAG's false discovery rate is alarmingly high in our setting.

We can additionally assess the credibility of the MTAG gains in power by evaluating whether the inflation of the mean  $\chi^2$ -statistic after applying MTAG are consistent with the increase in predictive power of a polygenic score (PGS) based on GWAS vs. MTAG summary statistics. If MTAG is leading to a spurious inflation of the  $\chi^2$ -statistics, the observed gains in predictive power will be smaller than the theoretical gains implied by the  $\chi^2$ -statistics.

The predictive power of a GWAS-based PGS<sup>15</sup> is

$$R^2 = \frac{h^4}{h^2 + \frac{M_{\text{eff}}}{N}},$$

where  $R^2$  is the predictive power of the GWAS-based PGS in the prediction sample,  $h^2$  is the heritability of the phenotype in the prediction sample,  $N$  is the size of the estimation sample used to create the PGS weights, and  $M_{\text{eff}}$  is the effective number of SNPs in the sample (assumed to be 30,000 here). We first use this relationship to infer the heritability of the phenotype in the prediction sample. More precisely, we substitute in the known values for  $R^2$ ,  $M_{\text{eff}}$ , and  $N$  and then solve for  $h^2$ .

To calculate the expected predictive power of the MTAG-based score, we replace  $N$  with the GWAS-equivalent sample size of the MTAG summary statistics,

$$N_{\text{equiv}} = N \frac{\overline{\chi_{\text{MTAG}}^2} - 1}{\chi_{\text{GWAS}}^2 - 1},$$

where  $\overline{\chi^2_{\text{MTAG}}}$  and  $\overline{\chi^2_{\text{GWAS}}}$  are the mean  $\chi^2$ -statistics of the MTAG and GWAS summary statistics, respectively.

Finally, we take the difference between the observed predictive power of the GWAS-based PGS and the expected predictive power of the MTAG-based PGS. We similarly calculate the difference between observed predictive power of the GWAS- and MTAG-based PGSs. If the inflation of the  $\chi^2$ -statistics in the MTAG results correspond to true gains in signal, these two differences should be similar.

A description of how the PGSs were constructed can be found in Section 6.2. The parameters for this derivation can all be found in **Supplementary Tables 41** and **42**. We find that the expected gains in the predictive power of the PGS for EA in Add Health is 0.4%, relative to an observed gain of 0.3% (CI  $-0.2\%$  to  $0.9\%$ ). For the PGS of EA in the HRS, we expect an increase of 0.4% and observe a gain of 0.6% (CI  $0.3\%$  to  $0.9\%$ ). For *CP*, in Add Health, we expect a gain of 1.9% and observe a gain of 1.8% (CI  $1.1\%$  to  $2.4\%$ ), and in the WLS, we expect a gain of 1.9% and we observe a gain of 2.7% (CI  $1.9\%$  to  $3.6\%$ ). In all cases, the expected gains are within the confidence intervals of the observed gains in predictive power, suggesting that the gains in  $\chi^2$ -statistics reported in the MTAG analyses represent a true increase in power.

## 2. Within-Family Association Analyses

### **2.1. Introduction**

The genotypes of full siblings are randomly assigned conditional on their biological parents' genotypes. Estimates from within-family association analyses are therefore immune to some types of confounding that can cause bias in association studies that rely, entirely or in part, on between-family variation. In this section, we report results from within-family association analyses conducted in a sample of  $N = 22,135$  sibling pairs. Given the small effect sizes of individual SNPs, our sibling sample is too small to allow well-powered within-family tests of association of single SNPs, including those with largest estimated effects; according to calculations in Rietveld et al.<sup>16</sup>, at least 47,000 pairs of siblings are needed for 80% power at  $P < 0.05$ . Instead, we test the SNPs jointly for association with *EduYears* in two different ways: we explore how often GWAS and within-family estimates have concordant signs, and we estimate the ratio of the true population parameters from within-family and GWAS analyses under different assumptions about the amount of stratification bias.

Below, we summarize methods used in the within-family association analyses and describe our subsequent implementation of the two tests (which we refer to as the sign test and the within-family regression test). To help interpret the findings, we also benchmark our results against predictions from a statistical framework that takes into account factors such as estimation error and winner's curse. Overall, the results in this section provide additional evidence in support of the conclusion that most of the observed increase in GWAS test statistics reflects polygenic signal. In order to make more quantitatively precise statements, it is necessary to make assumptions about the true ratio of within-family and GWAS parameters. For example, if we assume (i) a ratio of one and (ii) stratification biases of the magnitude implied by our estimated LD Score intercept, the within-family effects are 40% lower than predicted by our statistical framework. We show that assortative mating with respect to causal loci probably accounts for about one third of this deflation. We end the section by discussing and analyzing possible causes of the unexplained deflation. Several converging lines of suggest that much of the remaining deflation reflects omitted-variable biases (confounding) in the GWAS estimates that arises due to gene-by-environment correlation.

### **2.2. Within-Family Association Analyses**

*Within-Family Analyses.* Our within-family association analyses were conducted in four cohorts with sibling data: *STR-Twingene* ( $N = 2,727$ ), *STR-SALTY* ( $N = 707$ ), *UKB* ( $N = 17,097$ ), and *WLS* ( $N = 1,604$ ). In each of the three cohorts, we conducted within-

family association analyses by regressing the sibling difference in residualized *EduYears*,  $\tilde{y}_i$ , on the sibling difference in genotype:

$$\Delta\tilde{y}_i = \Delta g_{ij}\beta_{WF,j} + \Delta\epsilon_i \quad (2.1)$$

where  $i$  indexes a sibling pair,  $j$  indexes a SNP, and  $\Delta$  is the sibling-difference operator. *EduYears* was residualized using the same set of controls as the primary analyses and the same SNP-level and subject-level filters as in the main analyses (but in order for a sib pair to be included in the final analyses, data on both members of the pair need to satisfy the filters). We additionally restricted the analyses to SNPs with minor allele frequency above 5% in each of the sibling cohorts. We subsequently meta-analyze the cohort-level summary statistics using an inverse-variance weighting of cohort-level estimates of the parameter  $\beta_{WF,j}$ . In all within-family analyses, we standardize the phenotype using the standard deviation in the GWAS sample. The common standardization allows us to compare the effect size estimates from the within-family association analyses to those from the GWAS, to which we now turn.

### 2.3. Selection of SNPs Analyzed in Within-Family Tests

We used summary statistics from the within-family association analyses to conduct several tests, described in detail below. All require summary statistics from a discovery GWAS conducted in a sample independent of the sibling cohorts. We therefore reran the primary meta-analysis using exactly the same procedures as those described in **Supplementary Section 1**, but omitting the siblings included in the within-family analyses. Specifically, we excluded the results files from *STR-Twingene*, *STR-SALTY* and *WLS* from the main meta-analysis, and we reran the *UKB* association analyses in a discovery sample that omitted all *UKB* siblings from the estimation sample. After meta-analysis, we retained SNPs whose effects were estimated in a discovery sample of at least 700,000 individuals, whose minor allele is above 5%, and which are available in all three sibling cohorts. These restrictions leave us with 4,649,795 SNPs.

In some of the stratification tests, we further restricted our final analyses to subsets of approximately independent SNPs. In our sign tests (Section 2.7), we restricted our analyses to subsets of SNPs generated using our clumping algorithm (at  $P$  value thresholds ranging from  $5 \times 10^{-8}$ ,  $5 \times 10^{-5}$  and  $5 \times 10^{-3}$ ). We applied the clumping algorithm to  $P$  values calculated without any inflation adjustment, so the number of approximately independent SNPs are not perfectly comparable to those we report in the full meta-analysis. The number of approximately independent SNPs at the various thresholds is: 1,318 (at threshold  $5 \times 10^{-8}$ ), 4,594 ( $5 \times 10^{-5}$ ) and 14,670 ( $5 \times 10^{-3}$ ). For the within-family regression test (Section 2.8), we restricted our analyses to a subset of approximately

independent SNPs identified using PLINK’s pruning algorithm<sup>17</sup>, –indep-pairwise, with the following parameters: a window of 50 SNPs, a window shift of 5 SNPs, and a pairwise  $r^2$  threshold of 0.1. This leaves us with 82,609 SNPs.

## 2.4. Calculating a Theoretical Benchmark for Within-Family Association Results

Here, we describe the statistical framework we use to make predictions about the results we should expect from the two tests of within-family association reported below. The framework always adjusts for two factors that will produce differences between GWAS and within-family estimates: sampling variation (both in the GWAS and within-family association analyses) and winner’s curse in GWAS effect-size estimates (for SNPs selected on the basis of some  $P$  value threshold). We also extend the basic framework to adjust for coefficient differences driven by assortative mating.

## 2.5. Winner’s Curse Adjustment

We assume the phenotype and the genotype have both been standardized to have mean zero and unit variance. We assume that the effect sizes of all SNPs are drawn from the following mixture distribution:

$$\beta_j \sim \begin{cases} N(0, \tau^2) & \text{with probability } \pi \\ 0 & \text{otherwise,} \end{cases} \quad (2.2)$$

where  $\tau^2$  is the variance of non-null SNPs and  $\pi$  is the fraction of non-null SNPs in our data. By the Central Limit Theorem, we note that the estimation error of  $\hat{\beta}_j$  is approximately normally distributed. Letting  $n_j$  denote the size of the discovery sample in which the test of association for SNP  $j$  was conducted, the variance of the estimation error is  $\sigma_j^2 \approx 1/n_j$ . Since the estimation error and the causal effects are drawn independently,

$$\hat{\beta}_j \sim \begin{cases} N(0, \tau^2 + 1/n_j) & \text{with probability } \pi \\ N(0, 1/n_j) & \text{otherwise.} \end{cases} \quad (2.3)$$

We estimate the parameters  $\tau^2$  and  $\pi$  by maximum likelihood (see Supplementary Note p. 59 in Okbay et al.<sup>18</sup> for further details and a derivation of the likelihood function). Applied to the association statistics for the 4,649,795 SNPs in our meta-analysis with sibling cohorts excluded, we obtain  $(\hat{\tau}^2, \hat{\pi}) = (4.5 \times 10^{-6}, 0.64)$ . Given values for  $\tau^2$  and  $\pi$ , the posterior probability that SNP  $j$  with estimated effect size  $\hat{\beta}_j$  is non-null can be calculated using Bayes’ Rule:



$$p_{\hat{\beta},j} = \frac{\frac{\pi}{\sqrt{\sigma_j^2 + \tau^2}} \phi\left(\frac{\hat{\beta}_j}{\sqrt{\sigma_j^2 + \tau^2}}\right)}{\frac{1 - \pi}{\sigma_j} \phi\left(\frac{\hat{\beta}_j}{\sigma_j}\right) + \frac{\pi}{\sqrt{\sigma_j^2 + \tau^2}} \phi\left(\frac{\hat{\beta}_j}{\sqrt{\sigma_j^2 + \tau^2}}\right)}. \quad (2.4)$$

where  $\phi(\cdot)$  is a standard normal pdf. The posterior distribution of the effect size of each non-null SNP is:

$$(\beta_j | \hat{\beta}_j, \beta_j \neq 0) \sim N\left(\frac{\tau^2}{\sigma_j^2 + \tau^2} \hat{\beta}_j, \frac{\sigma_j^2 \tau^2}{\sigma_j^2 + \tau^2}\right). \quad (2.5)$$

For derivations of these results, we again refer to the Supplementary Note in Okbay et al.<sup>18</sup>.

For every SNP, it is straightforward to use the above equations to calculate the posterior probability that the SNP is non-null and its posterior effect-size distribution. To illustrate, consider a SNP whose effect is estimated to be  $\hat{\beta}_j = 0.005$  (implying  $R^2 = 0.0025\%$ ) in a discovery GWAS conducted in  $N = 750,000$  individuals. Substituting in the maximum likelihood estimates of  $(\hat{\tau}^2, \hat{\pi})$  and  $\sigma_j^2 \approx 1/n_j$  into the above equation for  $p_{\hat{\beta},j}$  yields probability  $p_{\hat{\beta},j} \approx 99.9\%$  that the SNP is non-null. And from the equation for the posterior distribution  $(\beta_j | \hat{\beta}_j, \beta_j \neq 0)$ , it is similarly straightforward to calculate the posterior mean of effect size distribution for the SNP. In our example,  $\frac{\hat{\tau}^2}{1/n_j + \hat{\tau}^2} \approx 0.75$ , so the GWAS effect estimate in this example would need to be shrunk by 25% to generate a winner's-curse-adjusted estimate of the SNP's effect (conditional on being non-null).

## 2.6. Calculating Theoretical Benchmarks

To illustrate how we use the framework to calculate theoretical benchmarks, let  $\hat{\beta}_{GWAS,j}$  denote the (standardized) GWAS estimate corresponding to SNP  $j$ , and let  $\hat{\beta}_{WF,j}$  denote the corresponding (standardized) estimate from a within-family association analysis. We denote their variances by  $\sigma_{GWAS}^2$  and  $\sigma_{WF}^2$ , respectively. We can decompose the estimates into orthogonal components as follow:

$$\hat{\beta}_{GWAS,j} = \beta_j + s_j + U_j \quad (2.6)$$

$$\hat{\beta}_{WF,j} = \beta_{WF,j} + V_j, \quad (2.7)$$

where  $s_j$  is the bias due to stratification,  $U_j$  and  $V_j$  are sampling errors with expected value 0, and  $\beta_j$  and  $\beta_{WF,j}$  are the true population parameters. Our theoretical benchmarks are always calculated for some fixed value of  $\text{Var}(s_j)$ , assuming that there is a single parameter ratio  $\beta_{WF,j}/\beta_j$  common to all SNPs. For example, in the sign tests, one of our theoretical benchmarks is calculated assuming  $\beta_j = \beta_{WF,j}$  (i.e., a ratio of one) and  $\text{Var}(s_j) = 0$ . In several of our analyses, we fix one of the two parameters ( $\text{Var}(s_j)$  or  $\beta_{WF,j}/\beta_j$ ) and calculate what the value of the second parameter would need to be to match our empirical findings. For example, in our sign tests, we show that if  $\text{Var}(s_j) = 0$ , then  $\beta_{WF,j}/\beta_j$  needs to be in the range 0.40 to 0.60 to match the realized sign concordance.

The assumption  $\beta_j = \beta_{WF,j}$  could be violated for a number of reasons, one of which is assortative mating with respect to genotypes at causal loci. Such assortative mating has been documented in analyses of genome-wide SNP data on spousal pairs and is substantial at both the genetic and phenotypic level. One recent paper reports a spousal phenotypic correlation of 0.41 for educational attainment in the UK Biobank<sup>19</sup>. In Section 2.10 below, we show that in a stylized model of phenotypic assortment<sup>20</sup>, it is possible to calculate analytically the attenuation of within-family coefficients relative to GWAS coefficients expected due to assortative mating. We show that for a trait with a large number of causal loci, within-family coefficients are deflated by a factor of  $[1 - rh_0^2]$ , where  $r$  is the spousal phenotypic correlation and  $h_0^2$  is the SNP heritability in a hypothetical base population without assortative mating. Accordingly, in several of our analyses below, we compare GWAS effect estimates to assortative-mating-adjusted within-family effect estimates. Specifically, we calculate the theoretical benchmark assuming that  $\beta_{WF,j}/\beta_j = R$  and  $\text{Var}(s_j) = 0$ , where  $R$  is an assortative-mating adjustment parameter. In those analyses, we set  $R = 0.83$ , a ballpark estimate based on published estimates of the relevant parameters for *EduYears*<sup>2,21</sup>.

Several of the large cohorts which contributed to our primary meta-analysis conducted their association analyses using mixed-linear models<sup>22,23</sup>. In these cohorts, listed in **Supplementary Table 19**, the adjustment for assortative mating is likely to be an overcorrection because SNPs were tested for association using models that attempt to control for the effects of all genetic variants that are in linkage equilibrium with the SNP. At least some of the linkage disequilibrium induced by assortative mating will therefore be controlled for in the mixed-linear-model effect estimates. We therefore interpret

results with and without the assortative-mating correction as bounds on the true parameter being estimated.

## 2.7. Sign Tests

Following Okbay et al.<sup>18</sup> we begin with a simple, transparent test that compares the sign concordance between the within-family (WF) estimates and GWAS estimates from a meta-analysis that excludes sibling cohorts.

*All Stratification.* Under the (strong) null hypothesis that the GWAS results are entirely driven by stratification, the sign of the WF estimates, which are immune to stratification, should be independent of the sign of the GWAS estimates and therefore should in expectation have a concordance of 50%. This means that among some set of  $M$  independent SNPs, the number of SNPs that have a concordant sign, denoted  $C$ , follows a binomial distribution:

$$C \sim \text{Binomial}(M, 0.5) \quad (2.8)$$

We can thus measure the observed sign concordance and use this known distribution to formally test the null hypothesis. We test the null hypothesis against the one-sided alternative hypothesis that the sign concordance is greater than 50%, which is what we would expect if at least some of the signal from the GWAS comes from true genetic effects. We conducted this test for sets of approximately independent SNPs selected at the  $P$  value thresholds  $5 \times 10^{-8}$ ,  $5 \times 10^{-5}$  and  $5 \times 10^{-3}$ .

*Winner's-Curse and Assortative-Mating Adjusted GWAS.* We also compare the observed sign concordance to the sign concordance expected when we set  $R = 1$  or  $R = 0.83$ . To simulate the distribution of the number of concordant signs under these two scenarios for a set of  $j = 1, 2, \dots, M$  approximately independent SNPs, we use the following procedure:

1. Draw a vector of true effect sizes from the posterior distributions, derived using the empirical-Bayes framework outlined in Section 2.5.
2. For each SNP, calculate the probability of sign concordance given its simulated, true effect size. As shown in Okbay et al.<sup>18</sup>, the probability of sign concordance for SNP  $j$  is:

$$p_j = \Phi\left(\frac{\beta_j}{\sigma_{\text{GWAS},j}}\right) \Phi\left(\frac{R\beta_j}{\sigma_{\text{WF},j}}\right) + \left[1 - \Phi\left(\frac{\beta_j}{\sigma_{\text{GWAS},j}}\right)\right] \left[1 - \Phi\left(\frac{R\beta_j}{\sigma_{\text{WF},j}}\right)\right], \quad (2.9)$$

where  $\Phi(\cdot)$  is the standard normal cumulative distribution function and  $\sigma_{GWAS,j}$  and  $\sigma_{WF,j}$  are the standard errors of the GWAS and within-family estimates. For these standard errors, we use  $\sigma_{GWAS,j} \approx \sqrt{1/n_j}$  and  $\sigma_{WF,j} \approx \frac{\sqrt{2(1-\rho)}}{\sqrt{n_{sibs,j}}}$ , where  $n_{sibs,j}$  is the number of sibling pairs used in the test of within-family association for SNP  $j$  and  $\rho$  is the sibling phenotypic correlation. The number of sign concordances is a sum of  $M$  independent success/failure experiments with success probabilities  $p_1, p_2, \dots, p_M$  that vary across SNPs. Hence, the total number of successes has a Poisson binomial distribution with parameter vector  $p = (p_1, p_2, \dots, p_M)$ .

3. For each simulation, we draw one realization of the total number of successes from the Poisson binomial distribution.

We repeat steps 1-3 10,000 times, each time recording the realized number of successes. Our final distribution assigns a probability of  $k$  successes that is equal to the average probability of  $k$  successes across the 10,000 iterations of the simulation. The expected sign concordance and its variance are the mean and variance of this distribution. Treating our estimate of the expected sign concordance as the probability that an independent SNP will have concordant sign, we conduct a one-sided binomial test where the alternative hypothesis is that our observed sign concordance is less than what we expect in the no-stratification case. We conducted this test for sets of approximately independent SNPs selected at the  $P$  value thresholds  $5 \times 10^{-8}$ ,  $5 \times 10^{-5}$ , and  $5 \times 10^{-3}$ .

*Results.* The results from our sign tests are shown graphically in **Figure 2.1** (see **Supplementary Table 20** for the underlying estimates). Across each of the  $P$  value cutoffs, the realized sign concordance is larger than would be expected if the results were entirely driven by stratification but smaller than predicted by our theoretical framework. For example, for the genome-wide significant SNPs ( $M = 1,318$ ), the observed concordance is 65.2%, 15.2 percentage points greater than the 50% we would expect if variance in GWAS estimates were entirely due to estimation error and bias, but also 6.6 percentage points lower than the theoretical benchmark calculated assuming  $\beta_{WF,j}/\beta_j = 0.83$  and  $\text{Var}(s_j) = 0$ . At cutoffs  $5 \times 10^{-5}$  and  $5 \times 10^{-3}$ , respectively, the observed concordance rates are 60.2% and 56.2%. Both realized concordances are approximately 6.5 percentage points lower than the theoretical benchmark. We also calculated the value of the ratio  $\beta_{WF,j}/\beta_j$  needed in order for the theoretical predictions to align with our findings under the maintained assumption that  $\text{Var}(s_j) = 0$ . We found that the ratio increases monotonically in the  $P$  value cutoff and ranges from 0.40 at  $P$  value cutoff  $5 \times 10^{-3}$  to 0.57 at cutoff  $5 \times 10^{-8}$ .

## 2.8. Within-Family Regression Test

Our next test also uses association results from a regular GWAS and a within-family association analysis. To motivate the within-family regression test, consider the population regression of  $\hat{\beta}_{WF,j}$  on  $\hat{\beta}_{GWAS,j}$ :

$$\begin{aligned} \frac{\text{Cov}(\hat{\beta}_{GWAS,j}, \hat{\beta}_{WF,j})}{\text{Var}(\hat{\beta}_{GWAS,j})} &= \frac{\text{Cov}(\beta_j + s_j + U_j, \beta_{WF,j} + V_j)}{\text{Var}(\beta_j + s_j + U_j)} \\ &= \frac{\text{Cov}(\beta_j, \beta_{WF,j})}{\text{Var}(\beta_j) + \text{Var}(s_j) + \sigma_{GWAS,j}^2}. \end{aligned} \quad (2.10)$$

Consider now an investigator wishing to make inferences about  $m_c \equiv \frac{\text{Cov}(\beta_j, \beta_{WF,j})}{\text{Var}(\beta_j) + \text{Var}(s_j)}$ . We will discuss below several special cases that clarify why this is an interesting parameter to estimate.

Since  $\text{Var}(\hat{\beta}_{GWAS,j}) = \text{Var}(\beta_j) + \text{Var}(s_j) + \sigma_{GWAS,j}^2$ , the analogy principle suggests the following estimator:

$$\widehat{m}_c \equiv \frac{\text{Cov}(\widehat{\beta}_{GWAS,j}, \widehat{\beta}_{WF,j})}{\text{Var}(\widehat{\beta}_{GWAS,j}) - \widehat{\sigma}_{GWAS}^2}. \quad (2.11)$$

(Note that this estimator is the slope from a regression of  $\widehat{\beta}_{WF,j}$  on  $\widehat{\beta}_{GWAS,j}$  with an adjustment term  $\widehat{\sigma}_{GWAS}^2$  that corrects the slope for estimation error in the GWAS summary statistics.) Our estimator for  $\widehat{\sigma}_{GWAS}^2$  is:

$$\widehat{\sigma}_{GWAS}^2 = \frac{1}{J} \sum_{j=1}^J \widehat{\sigma}_{GWAS,j}^2, \quad (2.12)$$

where  $J$  is the full number of SNPs in the GWAS (whereas  $M < J$  above referred to the number of SNPs after pruning to ensure approximate independence between the SNPs). We have that:

$$\text{E}[\widehat{m}_c] = \frac{\text{Cov}(\beta_j, \beta_{WF,j})}{\text{Var}(\beta_j) + \text{Var}(s_j)}. \quad (2.13)$$

Equation (2.13) shows that the expected value of the estimator depends on the amount of stratification in the GWAS estimates, the covariance of GWAS and within-family effects, and the variance of GWAS effect sizes.

To help interpret the estimator, it is instructive to begin by considering two special cases. A first arises when  $\beta_j = \beta_{WF,j}$ . In that case,  $\text{Cov}(\beta_j, \beta_{WF,j}) = \text{Cov}(\beta_j, \beta_j) = \text{Var}(\beta_j)$ . Then  $E[\widehat{m}_c]$  converges to  $\frac{\text{Var}(\beta_j)}{\text{Var}(\beta_j) + \text{Var}(s_j)}$ : the fraction of variance in the GWAS estimates that is due to true signal. In the extreme case in which the GWAS estimates capture no true signal and are entirely due to population stratification,  $E[\widehat{m}_c]$  will converge to zero. Under the null hypothesis that there is no stratification,  $E[\widehat{m}_c]$  will converge to 1. In contrast, in the presence of population stratification, the estimator will converge to some value less than one. Thus,  $\widehat{m}_c$  estimates can be directly compared to estimates from LD Score regression of the total amount of inflation in test statistics that is plausibly due to bias. (If instead  $\beta_{WF,j}/\beta_j = R$ , then the same conclusions hold but for  $R^{-1}\widehat{m}_c$  instead of  $\widehat{m}_c$ .) A second special case arises if  $\text{Var}(s_j) = 0$ , in which case we have:

$$E[\widehat{m}_c] = \frac{\text{Cov}(\beta_j, \beta_{WF,j})}{\text{Var}(\beta_j)}. \quad (2.14)$$

The expected value of the estimator is now the slope from the population regression of the GWAS coefficients on the within-family coefficients.

Following the convention introduced by Bulik-Sullivan et al.<sup>8</sup>, we estimate the standard errors using a block-jackknife approach, where blocks consist of sets of approximately 100 adjacent SNPs. More precisely, each SNP is assigned to one of 1000 equally-sized blocks, with block one consisting of the first 95 SNPs on chromosome 1, block two consisting of the next 95 SNPs, and so on through the whole genome. Then  $\widehat{m}_{c,b}$  is estimated 1,000 times, omitting a different block of SNPs each time. The standard error of  $\widehat{m}_c$  is then estimated to be

$$\text{SE}(\widehat{m}_c) = \left[ \frac{999}{1000} \sum_b (\widehat{m}_{c,b} - \widehat{m}_c)^2 \right]^{\frac{1}{2}}, \quad (2.15)$$

where  $b$  indexes the blocks.

Applying this method to the within-family estimates ( $\widehat{\beta}_{WF}$ ) derived from a meta-analysis of the three sibling cohorts ( $N = 22,135$  pairs) and GWAS estimates derived from a meta-analysis with the sibling cohorts omitted ( $N = 1,070,751$ ), we estimate  $\text{Cov}(\widehat{\beta}_{GWAS,j}, \widehat{\beta}_{WF,j}) = 1.28 \times 10^{-6}$ ,  $\text{Var}(\widehat{\beta}_{GWAS,j}) = 3.34 \times 10^{-6}$ , and  $\widehat{\sigma}_{GWAS}^2 = 9.44 \times 10^{-7}$ . Substituting these terms into the equation for  $\widehat{m}_c$  our lower-bound estimate ( $R = 1.0$ ) of  $m_c$  equals 0.53 and has a standard error of 0.04 (95% CI 0.46 to 0.60). Our

assortative-mating adjusted ( $R = 0.83$ ) upper-bound estimate is 0.64 and has a standard error of 0.05 (95% CI 0.55 to 0.72). Reassuringly, the cohort-level estimates are quite similar (**Supplementary Table 21** and **Supplementary Figure 21**).

According to the LD Score regression analyses in **Supplementary Section 1**, approximately 95% of the overall inflation in test statistics is due to polygenicity. Assuming this conclusion is correct, we can use equation (2.13) and our lower-bound estimate  $\widehat{m}_c = 0.53$  to estimate the ratio of within-family to GWAS parameters:

$$\frac{\text{Var}(\beta_j)}{\text{Var}(\beta_j) + \text{Var}(s_j)} = 0.95, \quad (2.16)$$

$$\frac{\text{Cov}(\beta_j, \beta_{WF,j})}{\text{Var}(\beta_j) + \text{Var}(s_j)} = 0.53. \quad (2.17)$$

Dividing (2.17) by (2.16) we have:  $\frac{\text{Cov}(\beta_j, \beta_{WF,j})}{\text{Var}(\beta_j)} = \frac{0.53}{0.95} = 0.56$  (or 0.67 if the assortative-mating adjustment is applied). Of course, the same exercise could be repeated under alternative assumptions about the share of inflation in GWAS test statistics that is due to bias.

## 2.9. Discussion and Additional Analyses

Summarizing the findings above, the results from both the sign tests and within-family regression test can be explained without assuming substantial amounts of stratification bias in GWAS estimates *only* if within-family parameters are ~45% smaller than GWAS parameters. Our calculations suggest that assortative mating is likely to cause deflation about one third as large, so assortative mating is likely to be an important but only partial explanation of the evidence. In this section, we report the results of a number of follow-up analyses that have two purposes: to help evaluate the robustness of our conclusions above and to provide some suggestive evidence relevant to evaluating possible explanations of the remaining difference between the GWAS and within-family estimates.

A first possibility is that data limitations or simplifying assumptions underlying our statistical model may impair the accuracy of its theoretical predictions. For example, our main analyses are based on parameter estimates inferred from summary statistics using analytic formulas that are subject to approximation error. Our assortative-mating correction is also derived assuming GWAS coefficients were estimated by OLS and is likely to overstate the amount of inflation in GWAS parameters for cohorts that used mixed linear models to conduct association analyses. Another concern is that the

statistical framework we use assumes a perfect genetic correlation between the GWAS and sibling samples. Phenotypic or genetic heterogeneity could cause this assumption to fail, in which case the theoretical benchmarks may be misleading.

In our first follow-up analysis, we performed the within-family regression test using only data from UK Biobank. We conducted the within-family association analyses in a sample of 17,097 pairs of siblings, and we compared the resulting estimates to estimates from a GWAS of UK Biobank respondents that excluded the siblings. Both association analyses were conducted using the same set of controls as in our primary GWAS of *EduYears*, and parameters were estimated by ordinary least squares (OLS). The primary goal of these analyses was to examine whether any of our main conclusions change when a phenotypically, genetically, and environmentally homogenous estimation sample is used. In this sample, we obtained an estimate of  $\widehat{m}_c = 0.36$  (SE = 0.06) before assortative-mating adjustment and 0.43 (SE = 0.08) after assortative-mating adjustment. Since the coefficient estimates were obtained by OLS, we do not consider the latter estimate an upper bound. The results from this first analysis provide little reason to believe that phenotypic, genetic, or environmental heterogeneity are major explanations for our findings.

Having established that our main result is unlikely to be an artefact explained by cohort heterogeneity, our next analyses, were designed to provide additional insights into possible causes of the lower-than-anticipated estimates of  $m_c$ . One possibility is that within-family parameters are deflated relative to GWAS parameters by environmental factors that reduce differences between siblings. For example, siblings may imitate each other, or parents may try to intervene to equalize sibling differences. A second possibility, also addressed in some of our analyses, is that the amount of stratification bias in GWAS estimates is greater than suggested by the LD Score regression results.

In our second analysis, we repeated the UKB-only within-family regression test, but using height as the phenotype instead of *EduYears*. Height provides a useful comparison because, a priori, it seems implausible that factors such as sibling interactions or compensatory investments by parents deflate  $\beta_{WF,j}$  relative to  $\beta_j$ . Therefore, if these factors are an important explanation for the low value of  $m_c$  found in our main analyses of *EduYears*, we expect a larger estimate when the methodology is applied to height. Our estimate,  $\widehat{m}_c = 0.90$  (SE = 0.02), is indeed substantially larger than the corresponding estimate for *EduYears*.

Third, we reran the first and second analyses of height and *EduYears* using a unified regression framework which generates two estimates per SNP, one based on between-family variation and one based on within-family variation (see Section 2.11 for details).



These analyses were based on an estimation sample limited to the UKB siblings and therefore has the additional advantage that the between-family and within-family estimates were estimated in the same set of individuals, using identical measures of the phenotype and genotype data. For height, our estimate is  $\widehat{m}_c = 0.70$  (SE = 0.03), and for *EduYears*, it is  $\widehat{m}_c = 0.40$  (SE = 0.04). Overall, the fact that these estimates are similar to those from our main analyses suggests that our main estimates are robust to alternative estimation strategies.

Our fourth analysis was designed to explore the possibility that  $\widehat{m}_c$  is low for *EduYears* because of subtle stratification biases that are more effectively handled by mixed linear association models than conventional OLS. (Our height results suggest that in order for this explanation to be plausible, the subtle stratification biases need to be specific to *EduYears*.) To test this possibility, we reran our first two analyses using association results from BOLT-LMM<sup>24</sup> instead of OLS. This possibility predicts that using BOLT-LMM would increase  $\widehat{m}_c$  by more than would be anticipated from the assortative-mating adjustment alone.

For height, the BOLT-LMM estimate is  $\widehat{m}_c = 0.98$  (SE = 0.02), compared to the original estimates of  $\widehat{m}_c = 0.70$  (SE = 0.03) (from the unified regression in the sibling sample) and  $\widehat{m}_c = 0.90$  (SE = 0.02) (OLS estimates in UKB with siblings omitted). Applying our assortative-mating adjustment increases the original estimates from 0.70 to 0.84 (unified regression) and from 0.90 to 1.08 (OLS). Therefore, the estimate of  $\widehat{m}_c = 0.98$  based on BOLT-LMM is approximately of the magnitude expected just from the assortative-mating adjustment. For *EduYears*, the BOLT-LMM estimate is  $\widehat{m}_c = 0.51$  (SE = 0.05), compared to the original estimates of  $\widehat{m}_c = 0.40$  (SE = 0.04) (unified regression in sibling sample) and  $\widehat{m}_c = 0.36$  (SE = 0.06) (OLS estimates in UKB cohort with siblings omitted). Adjusted for assortative mating, the original estimates are 0.48 and 0.43, so the BOLT-LMM estimate of  $\widehat{m}_c = 0.51$  is greater than expected from just the assortative mating adjustment, but only marginally so. Overall, these analyses suggest that biases handled more effectively by BOLT-LMM than OLS are not likely to be a major explanation for our lower-than-expected estimate of  $m_c$ .

The results from our follow-up analyses do not rule out stratification biases as a potentially important factor but limit the set of hypotheses about the types of stratification biases that are plausible. In particular, in order to fit the results, a plausible theory needs to posit a confounding variable that introduces substantial stratification bias (i) in association analyses of *EduYears* but not height, (ii) that are not effectively accounted for by mixed-linear association models (or OLS), and (iii) that evade detection by LD Score regression.

Evidence from behavior-genetic studies suggests that one unobserved variable that may satisfy these conditions is the individual’s rearing environment in childhood and adolescence<sup>25,26</sup>. Adoption studies have consistently found that children adopted into households with college-educated parents are more likely to attend college<sup>26–28</sup>. For example, a study of Korean-American adoptees finds that adoptees assigned to households where both parents had college degrees were 16 percentage points more likely to attend college than children assigned to families where neither parent completed college<sup>27</sup>. By contrast, the adult height of adoptive children is unrelated to the height of their adoptive parents<sup>27,29</sup>. Similarly, two genetically unrelated siblings reared in the same Swedish household are no more similar in height than two randomly chosen individuals, whereas the correlation in their *EduYears* is around 0.20<sup>29</sup>.

The evidence thus suggests that the effect of the rearing environment on *EduYears* is substantial, both in absolute terms and relative to many other phenotypes, and that the specific features of the rearing environment that matter for *EduYears* are likely to be strongly correlated with the *EduYears* of the parents in the rearing household. This observation, coupled with the fact that EA is heritable, implies we should expect a positive relationship between an individual’s EA genotype and their rearing environment in households with children raised by their biological parents. Several empirical studies have documented this positive relationship empirically using EA polygenic scores<sup>30,31</sup>. Thus, it is plausible that a GWAS of *EduYears* that fails to control for the rearing environment will yield upward-biased estimates, whereas a similar omission in an otherwise identical GWAS of height would be inconsequential. Intuitively, the source of the upward bias is that GWAS coefficient estimate will reflect both the causal effect of the SNP and its positive correlation with (the unobserved variable) rearing environment.

Though it is plausible that the magnitude of the gene-environment correlation will vary, if the rearing environment is highly correlated with parental EA, then there will be a systematic tendency for the correlation to be positive. This correlation will lead to a correlation between the stratification bias and the LD Score. As a result, this bias will not be reflected in the intercept of LD Score regression (see Supplementary Note 2.12).

A natural starting point for empirically testing the hypothesis proposed here would be to conduct association analyses with and without controls for parental phenotypes (using otherwise identical methods to those in our primary GWAS) and estimate the ratio of the GWAS estimates with and without the parental controls. If the proposed explanation is correct, adding parental controls would substantially reduce the ratio when the dependent variable is *EduYears*, but not when it is height. Evidence to the contrary would suggest that the omitted-variable bias posited here is not a quantitatively important explanation of the low  $\widehat{m}_c$  we report for *EduYears*.

In one of our prediction cohorts, HRS, information is available about mothers' and father's *EduYears* (but not height) for  $N = 7571$  European-ancestry respondents, allowing some exploratory genome-wide association analyses with and without controls for the midpoint of parental *EduYears*. All analyses were restricted to the SNPs used in our within-family regression tests and included the same baseline controls used in the cohort-level analyses for the primary GWAS meta-analysis. Let  $\hat{\beta}_{NP,j}$  and  $\hat{\beta}_{P,j}$  be the estimated effects of SNP  $j$  with and without parental controls, respectively. Procedurally, we obtained these estimates from a single regression in which each individual is entered as two observations. To account for the resulting non-independence, standard errors are clustered at the level of the individual<sup>32</sup>. The estimator is

$$\widehat{m}_c \equiv \frac{\text{Cov}(\widehat{\beta}_{NP,j}, \widehat{\beta}_{P,j}) - \widehat{\sigma}_{NP,P}^2}{\text{Var}(\widehat{\beta}_{NP,j}) - \widehat{\sigma}_{NP}^2}. \quad (2.18)$$

where  $\widehat{\sigma}_{NP}^2 = \frac{1}{J} \sum_{j=1}^J \widehat{\sigma}_{GWAS,j}^2$  and  $\widehat{\sigma}_{NP,P}^2 = \frac{1}{J} \sum_{j=1}^J \widehat{\sigma}_{NP,j,P,j}^2$  are straightforward to calculate (due the single-regression procedure) from the variance-covariance matrix of the estimates at each SNP. Our point estimate is  $\widehat{m}_c = 0.76$ , with a standard error of 0.14 (95% CI 0.48 to 1.03). Though this effect is in the hypothesized direction, it is not estimated precisely enough to permit any strong conclusions. Moreover, data limitations in HRS preclude us from repeating the same analysis with height instead of *EduYears*.

As this study was undergoing revisions, a paper with further corroborating evidence became available. Using a large sample of genotyped parent-child pairs from Iceland, the study documented that a polygenic score for *EduYears* constructed entirely from non-transmitted parental alleles predicts a respondent's educational attainment. A plausible interpretation of this finding is that non-transmitted alleles are associated with *EduYears* through their effects on the child's rearing environment. The effect of a polygenic score based on non-transmitted alleles was approximately 30% as large as the effect of a polygenic score based on transmitted alleles. An analogous analysis of height found that the effect of the non-transmitted-allele score was 6% as large as the effect of the transmitted allele score.

Overall, several converging lines of evidence suggest that in our setting, the LD Score intercept understates the share of inflation in GWAS test statistics that is due to bias. In behavior-genetic parlance, education is partly the result of vertical transmission of the parental phenotypes, which induces passive gene-by-environmental correlation that is not fully accounted for in association analyses. The source of bias conjectured here operates by amplifying a true underlying genetic effect and hence would not lead to false discoveries<sup>33</sup>. However, the environmental amplification implies that we should usually

expect GWAS coefficients to provide exaggerated estimates of the magnitude of causal effects. Such exaggeration implies that one must exercise care when interpreting genetic associations with phenotypes such as EduYears. For example, polygenic scores are often described as measures of genetic endowments. It is widely appreciated that they can influence complex outcomes by evoking certain environments, but the possibility that they are partly proxies for the environmental conditions under which a person was raised is rarely considered.

## 2.10 Appendix: Derivation of Adjustment for Assortative Mating

One reason that within-family estimates of SNP effects may be different than GWAS estimates based on samples of unrelated individuals is assortative mating. When parents' phenotypes are correlated, this will induce long-range LD across SNPs, even between chromosomes. Here we outline a very simple framework to approximate the difference in effect sizes we would expect to see as a result of assortative mating.

In this derivation, we make several simplifying assumptions. First, we assume that the only source of correlation between SNPs is assortative mating. That is, in a randomly mating population, causal SNPs would not be in LD. Second, we assume that all causal SNPs have the same effect on the phenotype. Third, we assume that, conditional on the genotypes of all other SNPs, the within- and between-family effect of each SNP is the same. This rules out sibling peer effects and other environmental effects that are endogenous to the child's phenotype. We anticipate that relaxing these assumptions would lead to the same substantive results.

We begin with a simple model:

$$y_i = \sum_{j \in C} x_{ij} \beta + \varepsilon_i,$$

where  $y_i$  is the measured phenotype for individual  $i$ ,  $C$  is the set of (independent) causal SNPs,  $x_{ij}$  is that individual's genotype,  $\beta$  is the common effect of each causal SNP on the phenotype, and  $\varepsilon_i$  is the residual, which we assume is independent of  $x_{ij}$ . In this model, we standardize the genotypes by dividing the allele count by  $\sqrt{2p_j(1-p_j)}$ , where  $p_j$  is the minor allele frequency of SNP  $j$ , such that the genotype will have variance one in the absence of assortative mating but may have variance greater than that if there is assortative mating in the population.

Because differences in sibling genotypes are assigned randomly within a family, within-family estimates of  $\beta$ ,  $\beta_{WF}$ , are unbiased by assortative mating. Therefore,  $\beta_{WF} = \beta$  for all SNPs. In contrast, standard GWAS estimates are based on between-family

comparisons. Therefore, the long-range LD induced by assortative mating will bias these estimates. Below, we derive the magnitude of this bias.

We begin by deriving the impact of assortative mating on the variance of each SNP and the covariance of each pair of SNPs. Let  $\tilde{x}_{ij} \in \{0,1,2\}$  denote the non-standardized genotype for individual  $i$ 's SNP  $j$ . This can be decomposed into a paternal and maternal component:  $\tilde{x}_{ij} = x_{ij}^{Pat} + x_{ij}^{Mat}$ . Assortative mating implies that paternal and maternal genotypes will be correlated. We denote this correlation by  $\alpha \equiv \text{Corr}(x_{ij}^{Pat}, x_{ij}^{Mat}) > 0$ . In fact, because we assume that all causal SNPs have the same effect size,  $\text{Corr}(x_{ij}^{Pat}, x_{ik}^{Mat}) = \alpha$  for any pair of causal SNPs  $j$  and  $k$ . We make the simplifying assumption that parents sort only based on  $y_i$  and other non-heritable factors, which will imply that parental genotypes are only correlated for causal SNPs.<sup>d</sup>

Since  $x_{ij}^{Pat}$  and  $x_{ij}^{Mat}$  are Bernoulli distributed with minor allele frequency  $p$ , we have

$$\begin{aligned}\text{Var}(\tilde{x}_{ij}) &= \text{Var}(x_{ij}^{Pat}) + \text{Var}(x_{ij}^{Mat}) + 2\text{Cov}(x_{ij}^{Pat}, x_{ij}^{Mat}) \\ &= 2p(1-p) + 2p(1-p)\alpha \\ &= 2p(1-p)(1+\alpha).\end{aligned}$$

Additionally, since we assume that all genotypes are in linkage equilibrium, different SNPs are inherited independently from the same parent and therefore  $\text{Cov}(x_{ij}^{Pat}, x_{ik}^{Pat}) = \text{Cov}(x_{ij}^{Mat}, x_{ik}^{Mat}) = 0$ . This implies that

$$\begin{aligned}\text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) &= \text{Cov}(x_{ij}^{Pat}, x_{ik}^{Pat}) + \text{Cov}(x_{ij}^{Mat}, x_{ik}^{Mat}) + \text{Cov}(x_{ij}^{Pat}, x_{ik}^{Mat}) \\ &\quad + \text{Cov}(x_{ij}^{Mat}, x_{ik}^{Pat}) \\ &= 4p(1-p)\alpha.\end{aligned}$$

The covariance and variance of the standardized genotypes,  $x_{ij}$ , is therefore inflated compared to the case of random mating:

$$\text{Var}(x_{ij}) = 1 + \alpha$$

---

<sup>d</sup> If we assumed otherwise, it is possible that assortative mating could inflate the Type I error rate slightly since a SNP that is associated with some phenotype that determines sorting, that SNP would be in long-range LD with all SNPs associated with the phenotypes that determine sorting. We anticipate that this correlation would be low relative to the long-range LD between SNPs associated with the same phenotype due to imperfect genetic correlation, but examining this issue more completely is beyond the scope of this paper.

$$\text{Cov}(x_{ij}, x_{ik}) = 2\alpha.$$

ote that these equations only hold for causal SNPs. The second equation therefore implies that assortative mating will induce a (potentially long-range) correlation between a pair of SNPs if both SNPs are truly associated with the phenotype.

Crow and Kimura<sup>34</sup> derived an expression for the size of the spousal SNP correlation for a causal SNPs,

$$\alpha = \frac{rh_0^2}{rh_0^2 + 2M(1 - rh_0^2)},$$

where  $r$  is the phenotypic correlation between spouses,  $h_0^2$  is the narrow-sense heritability under random mating, and  $M$  is the number of causal loci. Using the above results, we show that the between-family GWAS coefficient for some causal SNP  $j$  under assortative mating is

$$\begin{aligned} \beta_{\text{BF},j} &= \frac{\text{Cov}(x_{ij}, y_i)}{\text{Var}(x_{ij})} \\ &= \frac{\text{Cov}(x_{ij}, \sum_{k \in C} x_{ik}\beta + \varepsilon_i)}{\text{Var}(x_{ij})} \\ &= \beta \frac{\sum_{k \in C} \text{Cov}(x_{ij}, x_{ik})}{\text{Var}(x_{ij})} \\ &= \beta_{\text{WF}} \frac{1 + (2M - 1)\alpha}{1 + \alpha}. \end{aligned}$$

Note that if SNP  $j$  weren't a causal SNP or in LD with a causal SNP, then  $\text{Cov}(x_{ij}, x_{ik}) = 0$  for all SNPs in the set  $C$ . So  $\beta_{\text{BF},j} = 0$ . This implies that assortative mating does not inflates the effect size estimates for null SNPs and therefore should not lead to inflated type-I error rates.

Returning to the case where SNP  $j$  is causal, substituting in Crow and Kimura's expression for  $\alpha$ , rearranging, and taking the limit as the number of loci goes to infinity, the ratio of within-family and between-family parameters converges to:

$$\lim_{M \rightarrow \infty} \frac{\beta_{\text{WF},j}}{\beta_{\text{BF},j}} = 1 - rh_0^2.$$

Thus, the ratio of within- and between-family estimates is (approximately) a simple function of the spousal correlation of the phenotype and the narrow-sense heritability under random mating. The only remaining unknown parameter is the narrow-sense

heritability under random mating,  $h_0^2$ . Crow and Kimura<sup>34</sup> derive the relationship between  $h_0^2$  and the narrow-sense heritability under assortative mating  $h^2$  as

$$h_0^2 = \frac{h^2 - h^4 r}{1 - h^4 r}.$$

We can thus obtain an estimate of  $h_0^2$  by plugging into this equation estimated values for  $r$  and  $h^2$  from the literature.

## 2.11 Appendix: Unified Regression Analyses

In these analyses, we continue to use equation (2.11) to estimate  $m_c$  but this time using coefficient estimates from the unified regression analyses. Specifically, for each SNP  $j$  our estimating equation is:

$$y_{if,j} = \beta_{0,j} + \beta_{GWAS,j} \bar{g}_{f,j} + \beta_{WF,j} (g_{if,j} - \bar{g}_{f,j}) + x_{if,j} \gamma_j + u_f + \varepsilon_{if,j},$$

where  $i$  indexes an individual,  $f$  indexes a family,  $y_{if,j}$ ,  $g_{if,j}$ , and  $x_{if,j}$  are the individual-specific phenotype, genotype, and vector of controls, respectively,  $\bar{g}_{f,j}$  is the within-family mean of the genotype,  $\varepsilon_{if,j}$  is an individual-specific error term, and  $u_f$  is a family-specific error term. For each SNP, the unified regression recovers two SNP effects. The first,  $\hat{\beta}_{GWAS}$ , is identified solely from between-family differences, and the second,  $\hat{\beta}_{WF}$ , is identified solely from within-family variation. We restrict the estimation sample to UKB siblings and all analyses are based on a phenotype that has been residualized on the set of baseline covariates used in all other within-family association analyses. The estimates of  $m_c$  reported are based on a set of 247,915 approximately independent SNPs.

## 2.12 Appendix: LD Score Regression and Childhood Rearing Environment

Here we show that when an individual's EA is influenced by parental EA, this will lead to bias in GWAS estimates that is not captured by the LD Score intercept. We assume a simple model:

$$y_i = \sum_j \beta_j g_{i,j} + \alpha_f y_i^f + \alpha_m y_i^m + \varepsilon_i,$$

where  $y_i$  is the standardized EA of individual  $i$ ,  $g_{i,j}$  is the standardized genotype of SNP  $j$  for individual  $i$ ,  $y_i^f$  and  $y_i^m$  are the standardized EA of the father and mother, respectively, and  $\varepsilon_i$  is the residual.

We make a number of simplifying assumptions. First, we assume that  $\varepsilon_i$  is independent of  $g_{i,j}$  so that parental educational attainment is the only source of

confounding in the model. If we modeled other sources of confounding, those sources may be reflected in the LD Score regression slope if the confounding results in bias that is proportional to the LD Score (as with the example shown in this section) or in the LD Score regression intercept if the confounding generates bias that is uncorrelated with the LD score (e.g., population stratification due to drift).

Second, we assume that the genetic factors that determine each parent's EA are the same as those that affect the child's EA and that the parents' rearing environment does not affect their EA. More precisely, we assume

$$y_i^p = \sum_j \beta_j g_{i,j}^p + \varepsilon_i^p,$$

where  $p \in \{f, m\}$  identifies the father or mother,  $g_{i,j}^p$  is the genotype of parent  $p$  of individual  $i$  at SNP  $j$ , and  $\varepsilon_i^p$  is the residual. Were we to assume that a parent's EA was similarly affected by their rearing environment, that would lead to a simple geometric series in the derivation below, resulting in the same qualitative results.

Third, we assume random mating, which implies that the genotypes of each parent are uncorrelated with each other. Thus,  $\text{Cov}(g_{i,k}, g_{i,j}^p) = \frac{1}{2} r_{j,k}$ , where  $r_{j,k} \equiv \text{Cov}(g_{i,k}, g_{i,j})$ . Even under assortative mating, however, this equation would hold approximately as long as the traits that determine how parents sort are polygenic. (See section 2.10 above.)

Under these assumptions, we derive the expected value of the GWAS estimates. In our GWAS, we regress  $y_i$  on each SNP separately. Since we standardized the genotypes, the expected GWAS estimate for SNP  $k$  is

$$\begin{aligned} E(\beta_{\text{GWAS},k}) &= \text{Cov}(g_{i,k}, y_i) \\ &= \text{Cov}\left(g_{i,k}, \sum_j \beta_j g_{i,j} + \alpha_f y_i^f + \alpha_m y_i^m + \varepsilon_i\right) \\ &= \text{Cov}\left[g_{i,k}, \sum_j (\beta_j g_{i,j} + \alpha_f \beta_j g_{i,j}^f + \alpha_m \beta_j g_{i,j}^m) + \varepsilon_i + \alpha_f \varepsilon_i^f + \alpha_m \varepsilon_i^m\right] \\ &= \sum_j \beta_j [\text{Cov}(g_{i,k}, g_{i,j}) + \alpha_f \text{Cov}(g_{i,k}, g_{i,j}^f) + \alpha_m \text{Cov}(g_{i,k}, g_{i,j}^m)] \\ &= \sum_j \beta_j \left(r_{j,k} + \frac{\alpha_f}{2} r_{j,k} + \frac{\alpha_m}{2} r_{j,k}\right) \end{aligned}$$



$$= \sum_j \beta_j \left(1 + \frac{\alpha_f}{2} + \frac{\alpha_m}{2}\right) r_{j,k}.$$

From this expression, we see that the bias due to the rearing environment is multiplicative. It follows that the GWAS effect sizes will be inflated, but the type-I error will not be inflated, as similarly shown by Lee and Chow<sup>35</sup>.

The fact that the bias due to the rearing environment is multiplicative also implies that LD Score regression cannot disentangle this source of bias from the true causal effect of the SNP. Indeed, it is simple to use the above expression for  $E(\beta_{\text{GWAS},k})$  to extend the derivation of the LD Score regression estimating equation, giving us

$$E(\chi_k^2) = \frac{N}{M} h^2 \left(1 + \frac{\alpha_f}{2} + \frac{\alpha_m}{2}\right)^2 \ell_k + 1,$$

where  $\ell_k$  is the LD Score of SNP  $k$ . So we see in this case that although the GWAS coefficients are inflated by a factor  $\left(1 + \frac{\alpha_f}{2} + \frac{\alpha_m}{2}\right)$ , this bias is not reflected in the intercept but rather in an inflated slope.

### 3. Heritability and Genetic Correlation Across Cohorts

#### **3.1 Introduction**

Genetic effects on *EduYears* may vary across cohorts due to, for example, environmental differences in the requirements for and motivations to progress through educational systems. Indeed, in **Supplementary Section 1**, we found that the distribution of  $P$  values among top SNPs deviates from that expected if genetic effects on *EduYears* are homogeneous across cohorts. This heterogeneity in genetic effects may reflect differences across cohorts in the heritability of *EduYears* or imperfect genetic correlation across pairs of cohorts. In this section, we estimate the amount of variation in SNP heritability of *EduYears* across cohorts in our sample and the average genetic correlation of *EduYears* across pairs of cohorts. In an exploratory analysis, we also assess the extent to which we can predict variation in SNP heritability and genetic correlation of *EduYears* based on several observable cohort characteristics. Hereafter in this section, for simplicity, we refer to SNP heritability as “heritability.”

In addition to being of substantive interest, variation in heritability and imperfect genetic correlation matters for how much predictive power we expect to obtain from a polygenic score (PGS) as a function of the characteristics of a prediction cohort. As partial motivation for the analysis, we therefore begin this section by briefly reviewing how heritability and genetic correlation are theoretically expected to affect predictive power of a PGS.

We then turn to the empirical analyses, which are the main focus of the section. First, we estimate the amount of variation in heritability across our cohorts as well as the mean genetic correlation of *EduYears* across pairs of cohorts. We estimate a standard deviation of heritability of 0.023 (SE = 0.048), compared with a mean of 0.147 (SE = 0.009). We also find that the mean genetic correlation of *EduYears* estimated across pairs of cohorts is 0.723 (SE = 0.124), which is statistically distinguishable from one ( $P$  value = 0.026).

These results motivate our remaining analyses, which explore the extent to which the variation in heritability and the imperfect genetic correlation across cohorts can be explained by several cohort characteristics that we examine. In a complementary set of analyses, we study heritability and genetic correlation of *EduYears* using individual-level data from a single, large cohort: the full-release UK Biobank. This analysis has the advantage that it holds constant many potentially confounding country-level variables that we cannot control for in the cross-cohort analysis, but it has the disadvantage that we cannot examine cohort characteristics that vary little within the UK. To preview the

results, the only observable cohort characteristic we examine that robustly explains variation in heritability is the level of detail of the survey assessment of *EduYears*. We also find suggestive evidence that the genetic correlation of *EduYears* is highest for cohorts with the same average birth year, but this result does not replicate in the UK Biobank analysis.

The remainder of this section is organized as follows. We begin by reviewing why the predictive power of a PGS may vary across cohorts as a function of heritability and genetic correlation (Section 3.2). In Section 3.3, we present our methods for and results from estimating heritability and genetic correlation across cohorts. In Section 3.4, we describe the cohort characteristics that we study which might influence heritability or genetic correlation. These include the number of response options for *EduYears* provided in a cohort, the average year of birth among cohort respondents, the average educational attainment in the cohort’s country, and top income shares in the country. In Sections 3.5 and 3.6, respectively, we present our exploratory analyses of the association of these cohort characteristics with heritability and genetic correlation of *EduYears*. In Section 3.7, we describe and report results from the UK Biobank analysis, in which we evaluate whether heritability and genetic correlation of *EduYears* differs depending on the survey question used to assess educational attainment or on respondents’ years of birth. We conclude with a discussion of the implications of our results for the predictive power of PGSs in Section 3.8.

### 3.2 How Population Differences Affect Cross-Cohort Prediction Accuracy

Daetwyler et al.<sup>15</sup> introduced the following formula for the expected predictive power of a PGS:

$$E[R^2] = h^2 \times \left[ \frac{1}{1 + \frac{1}{\lambda h^2}} \right], \quad (3.1)$$

where  $R^2$  is the coefficient of determination in the prediction cohort,  $h^2$  is the SNP heritability of the phenotype, and  $\lambda \equiv n_D/M$  is the ratio between the number of individuals in the discovery cohort ( $n_D$ ) and the effective number of SNPs evaluated in the prediction cohort ( $M$ )<sup>e</sup>. The formula is derived under the following three key assumptions: (1) the SNP genotypes are independent, (2) the heritability of the phenotype

---

<sup>e</sup> The value of  $M$ , which reflects the amount of *independent* variation in genotypes across SNPs (which is smaller than the number of SNPs due to LD), differs depending on the set of SNPs that are used. For example, for HapMap3 SNPs,  $M$  is likely in the range 50,000 to 70,000<sup>334</sup>.

in the discovery and prediction cohorts is equal, and (3) across the cohorts, there is perfect genetic correlation of the phenotype.

De Vlaming et al.<sup>36</sup> generalized the formula, relaxing the latter two assumptions (but maintaining the first). That is, the more general formula applies when the discovery and prediction cohorts have unequal heritability ( $h_D^2$  and  $h_P^2$ , respectively) and imperfect genetic correlation ( $r_g < 1$ ):

$$E[R^2] = r_g^2 \times h_P^2 \times \left[ \frac{1}{1 + \frac{1}{\lambda h_D^2}} \right]. \quad (3.2)$$

The formula implies that the expected predictive power of the PGS increases with increasing values of  $h_D^2$ ,  $h_P^2$ , and  $r_g$ . Thus, we expect a PGS to have lower  $R^2$  in a prediction cohort to the extent that the phenotype measured in the prediction cohort is imperfectly genetically correlated with the phenotype in the discovery cohort, has a lower heritability in the discovery cohort, and has a lower heritability in the prediction cohort.

### 3.3 Variation in Heritability and Mean Genetic Correlation of *EduYears* Across Cohorts

Heritability of *EduYears* could vary across cohorts for many reasons, including differences in the accuracy with which the phenotype is measured and differences in the institutional environment (e.g., social and economic circumstances) experienced by respondents. Relatedly, *EduYears* may be imperfectly genetically correlated across cohorts if the effects of genetically influenced psychological characteristics (e.g., specific cognitive abilities or personality traits such as persistence) on *EduYears* differ. This could be due to differences in the educational systems or labor markets, for example. In this subsection, we calculate variation in heritability across cohorts and estimate the mean genetic correlation of *EduYears* for pairs of cohorts in our sample.

#### Empirical methods

We first estimate the heritability of *EduYears* in each of the 71 cohorts included in the meta-analysis using LD Score regression<sup>8</sup>, as implemented in the LDSC software with a European reference population. LDSC filters to HapMap3 SNPs because these SNPs are generally well imputed and because LD Scores of SNPs with low minor allele

frequencies are known to introduce substantial statistical noise into the analysis<sup>f</sup>. **Supplementary Table 22** reports the heritability estimates for each cohort.

Unfortunately, the estimated heritability of *EduYears* is negative for 23 of the 71 cohorts, and the genetic correlation as calculated by LD Score regression<sup>14</sup> is undefined when one cohort is estimated to have a negative heritability. The negative estimates appear to be primarily due to sampling error in smaller cohorts, as standard errors for the negative heritability estimates are generally large, and the majority of their 95% confidence intervals contain 0. Fortunately, the 32% of cohorts with negative heritability estimates represent only 6% of respondents. We therefore proceed with our analysis limiting our data to the 48 cohorts (containing 1,060,743 respondents) with non-negative heritability estimates.

We calculate the mean heritability of *EduYears* by taking the sample-size-weighted mean of the estimates across our 48 cohorts. We then estimate the cross-cohort variance in heritability. Note that we are interested in the cross-cohort variance in *true* heritability. We cannot simply estimate that quantity with the variance in our heritability estimates because variation in the heritability estimates reflects sampling variation in addition to true variation. Instead, to obtain our estimate of the cross-cohort variance in heritability, we use a regression framework, explained next.

As a starting point, note that if our estimator of heritability is unbiased and its sampling variance scales linearly with the inverse of the sample size—as is approximately true for our LD Score regression estimator—then  $E(\hat{h}_c^2 | h_c^2) = h_c^2$  and  $\text{Var}(\hat{h}_c^2 | h_c^2) = \lambda/N_c$ , where  $h_c^2$  is the heritability for the cohort  $c$ ,  $\hat{h}_c^2$  is its estimate,  $N_c$  is the cohort sample size, and  $\lambda$  is some constant that does not vary by cohort. Therefore, using  $\sigma_h^2 \equiv \text{Var}(h_c^2)$  to denote the variance of heritability across cohorts (our parameter of interest), we have

$$\text{Var}(\hat{h}_c^2) = \frac{1}{N_c} \lambda + \sigma_h^2. \quad (3.3)$$

Our estimator for the variance of heritability is based on this equation: if we knew  $E(h_c^2)$ , we could regress  $[\hat{h}_c^2 - E(h_c^2)]^2$  onto  $N_c^{-1}$  and a constant, and the intercept would be an estimate of the true variance of heritability across cohorts,  $\sigma_h^2$ . Since  $E(h_c^2)$  is

---

<sup>f</sup> Additional filters were applied to the GWAS data in its quality-control stage (see Section 1.6): SNPs with minor allele frequencies below 0.05 were excluded, as were those with imputation quality below 0.3.

unknown, in practice, we replace it with the sample-size-weighted mean described above. We then use weighted least squares, weighting each cohort by its sample size to obtain an estimate,  $\hat{\sigma}_h^2$ .

Turning to genetic correlation, we estimate the genetic correlation of *EduYears* across all unique pairs of cohorts with non-negative heritability estimates. To do so, we use bivariate LD Score regression<sup>14</sup> implemented by the LDSC software with a European reference population, filtered to HapMap3 SNPs. The estimated genetic correlation of *EduYears* between each of our 933 pairs of cohorts is shown in **Supplementary Table 23**.

We obtain the mean genetic correlation by estimating the inverse-variance-weighted mean of the estimates across pairs of cohorts in our data. The genetic correlation of *EduYears* across pairs of cohorts will be correlated across all observations that share one of their cohorts in common. Therefore, to obtain correct standard errors, we use the node-jackknife variance estimator described by Cameron and Miller<sup>37</sup>.

## Results

The sample-size-weighted mean heritability of *EduYears* across cohorts is 0.147 (SE = 0.009). We estimate the variance to be 0.00051, which implies a standard deviation estimate of  $\sqrt{0.00051} \approx 0.023$ . The variance is quite imprecisely estimated, however, with a standard error of 0.0023, which implies that the standard error of the standard deviation of heritability is approximately  $\sqrt{0.0023} \approx 0.048$ . This indicates that we are not well-powered to detect variation in heritability across cohorts.

The inverse-variance-weighted mean genetic correlation<sup>§</sup> of *EduYears* between pairs of cohorts is 0.723 (SE = 0.124). The *P* value for the null hypothesis of perfect genetic correlation is 0.026. This result reinforces the conclusion that there is some heterogeneity in genetic effects on *EduYears* across the 48 cohorts studied.

### 3.4 Observed Cohort Characteristics

In Sections 3.5 and 3.6, we will explore empirically the extent to which we can predict variation in heritability and genetic correlation across the cohorts in our sample based on several observed cohort characteristics. The four cohort characteristics that we identified reflect either the accuracy with which *EduYears* was measured or the

---

<sup>§</sup> For analyses of cross-cohort genetic correlation, we use inverse-variance weighting rather than sample-size weighting. We do this because there is not a simple relationship between the sampling variance of the genetic correlation estimate and the sample size of the two cohorts involved. Inverse-variance weights are calculated using the standard errors of the genetic correlation estimates reported by the LD Score regression software.

institutional environment experienced by individuals in the cohort.<sup>h</sup> Here, we describe the rationale underlying the selection of each of the four variables and the methods for their construction across our 48 cohorts.

*Number of response options for EduYears in the cohort:* This variable indicates the number of response options that were available to respondents of a particular cohort when reporting their educational attainment. If the question used to measure a phenotype provides few response options, the phenotype will be measured less accurately. Measurement error should in turn reduce the estimated heritability of the phenotype. Therefore, we expect the estimated heritability of *EduYears* to be smaller within cohorts for which fewer response options were provided. We do not expect this variable to have any effect on genetic correlation of *EduYears* between cohorts.

Across the 48 cohorts we study, the granularity of the question used to measure *EduYears* varies substantially: some cohorts provided just four response options, while others provided as many as 20 (see **Supplementary Table 22** and **Supplementary Table 24**, Panel A). The mean number of response options for reporting *EduYears* across cohorts, weighted by sample size, is 7.7.

*Mean birth year in the cohort:* Prior research using twin studies suggests that the heritability of *EduYears* is greater among people born in more recent years (e.g. Branigan et al.<sup>25</sup>), a trend that is hypothesized to be related to the expansion of access to education. In accordance with this literature, we expect that the heritability of *EduYears* as measured from the GWAS results (using LD Score regression Bulik-Sullivan et al.<sup>8</sup>) will be higher in cohorts with a later mean birth year.

We also expect cohorts that differ in mean birth year to evince imperfect genetic correlation of *EduYears*. The effects of certain genetically influenced psychological traits on *EduYears* may depend on the institutional and economic environment, which affects the motivations for and skills required to advance through education. Changes in this environment will result in imperfect genetic correlation of *EduYears* for people born in different historical periods.

---

<sup>h</sup> We selected these four characteristics because (i) we believed they might plausibly matter for heritability or genetic correlation, (ii) we were able to obtain measures of these characteristics for all cohorts, and (iii) we confirmed that they varied substantially across cohorts. There are several additional (or alternative) cohort characteristics that we could have assessed, but to mitigate concerns about multiple hypothesis testing and to maximize sample size, we did not pursue these. Some of the other cohort characteristics that we considered include the country which cohort respondents are primarily from and the cohort's sex composition (following Branigan et al.<sup>25</sup>); the compulsory education laws in the cohort's country, which could be a constraint on the expression of genetic propensity for EA; and the proportion of the labor force employed in manufacturing as a proxy for the availability of low-skill jobs, which might be an important determinant of the motivations for schooling.

As shown in **Supplementary Table 22** and **Supplementary Table 24**, Panel A, the mean year of birth among cohort respondents ranges from 1921 to 1979, with a sample-size-weighted average of 1954. Note that, although there is also substantial variation in birth years within cohorts, we are unable to exploit this variation because we do not have access to individual-level data for most of the cohorts in the analysis. An important exception is the UK Biobank cohort, and we do examine individual-level variation in birth year in that cohort in Section 3.7 below.

*Mean years of education in the cohort's country:* There are several channels through which mean level of education might be related to heritability of *EduYears*, although the direction of the relationship could go either way. For example, mean years of education in a country might proxy for access to education, which could increase heritability (just as in the argument above, where mean birth year proxied for access to education). On the other hand, mean years of education in a country might proxy for institutions (e.g., compulsory schooling laws or social norms prescribing college attendance) that compel people to attain a level of education that they otherwise would not. This might compress variation in *EduYears*, thereby reducing heritability.

There may also be imperfect genetic correlation of *EduYears* between cohorts with different mean years of education. For example, mean years of education might be correlated with labor market opportunities, and the decision to continue with schooling may be affected by different genetically influenced psychological characteristics in different labor markets.

For every country represented by a cohort, we obtained a measure of mean years of education among those ages 25 and above in 1950 (or the earliest available year) from a previously published study<sup>38</sup>. As reported in **Supplementary Table 22** and **Supplementary Table 24**, Panel A, mean educational attainment in a cohort's country in 1950 ranges from 3.8 years in Greece to 8.7 years in Switzerland; the sample-size-weighted mean across cohorts is 6.9 years.

*Top 10% income share in the cohort's country:* Similar to mean educational attainment in a country, heightened income inequality in a country may be associated with weaker heritability of *EduYears* if inequality reflects constraints on educational attainment for some parts of the population, compressing the distribution of *EduYears*. On the other hand, greater income inequality may reflect steeper returns to education, sharpening the incentives for individuals predisposed to further education to become more educated, which could result in a greater heritability of *EduYears*.

In populations with different levels of income inequality, we might also expect the genetic correlation of *EduYears* to be less than one. For example, in a country with



minimal income inequality, the decision to attain a high level of education could be driven to greater extent by genetically influenced personality traits associated with enjoyment of schooling, whereas in a less equal society, schooling decisions might be affected more by genetically influenced personality traits associated with status-seeking.

To measure income inequality, we use the share of income earned by the top 10% of earners in the country. This data is obtained from the World Wealth and Incomes Database (<http://www.wid.world>). When data was not available from this source, we use information from the World Bank Development Indicators database (<http://data.worldbank.org/data-catalog/world-development-indicators>). Where possible, we use income inequality data for the mean birth year of the cohort; otherwise, we use the available year closest to the mean birth year. Our income-inequality variable ranges from 18% for an Estonian cohort to a high of around 45% in some cohorts from the USA and the Netherlands; the sample-size-weighted mean is 31.4% (see **Supplementary Table 22** and **Supplementary Table 24**, Panel A).

As shown in **Supplementary Table 24**, Panel B, the sample-size-weighted Pearson correlations between the variables across all 48 cohorts included in this analysis range from  $-0.23$  to  $0.61$ . **Supplementary Table 22** provides further details about each cohort, including its country, sample size, and values for each of our four variables.

### 3.5 Cohort Characteristics and Heritability of *EduYears*

#### Empirical methods

To study the association of specific cohort characteristics with the heritability of *EduYears*, we use weighted linear regression with sample-size weighting. The dependent variable is the estimated heritability for each cohort (methods for estimating heritability are described in Section 3.3). We include as independent variables the four continuous measures of cohort characteristics described in Section 3.4, centered at their sample-size-weighted means.<sup>i</sup>

---

<sup>i</sup> In a preliminary analysis before the availability of the full UK Biobank (and before the addition of a few other cohorts), we pursued an alternative approach. The goal of this approach was to avoid excluding cohorts with negative heritabilities from our exploratory analyses of the association of cohort characteristics with heritability and genetic correlation. In this alternative analysis, we dichotomized each of the cohort characteristics that we study (described in Section 3.4) at their sample-size-weighted medians, and grouped cohorts into “profiles” depending on their values for each of these four binary variables. Nine profiles were non-empty. We then meta-analyzed the GWAS summary statistics of all cohorts that shared the same profile. We proceeded with the profiles as the unit of analysis rather than the individual cohorts, estimating the heritability of *EduYears* in each of the 9 profiles and estimating genetic correlation of *EduYears* for each of the  $\binom{9}{2} = 36$  pairs of profiles. Using the dichotomized variables in regression analyses of heritability and of genetic correlation, our results were qualitatively similar to those we find in

## Results

**Supplementary Table 25**, Panel A, shows the results from our regression analysis of heritability. The intercept indicates that, for cohorts with mean values on all four variables, the expected heritability of *EduYears* is 0.147 (significantly distinguishable from zero with  $SE = 0.008$  and  $P$  value =  $3.44 \times 10^{-22}$ ). As expected, the heritability of *EduYears* is significantly higher for cohorts that provided a greater number of response options ( $\beta = 0.006$ ;  $SE = 0.002$ ;  $P$  value = 0.009). The heritability of *EduYears* is expected to increase by a little over half of a percentage point with each additional response option provided. Results may also indicate higher heritability of *EduYears* in cohorts from countries with larger top income shares ( $\beta = 0.005$ ;  $SE = 0.002$ ;  $P$  value = 0.038). For the other two variables, we cannot statistically distinguish regression coefficients from zero (mean birth year  $\beta = 0.001$ ,  $SE = 0.002$ ,  $P$  value = 0.395; average education in the country  $\beta = -0.009$ ,  $SE = 0.011$ ,  $P$  value = 0.407).

Together, these four cohort characteristics explain 20.7% of variation in heritability estimates across cohorts, as indicated by the  $R^2$ . Much of the residual variation in the dependent variable is likely to be sampling error since the heritabilities are estimated, rather than fixed, known quantities. Thus, the  $R^2$  value may be considered a lower bound on the role of these four cohort characteristics in explaining actual cross-cohort variation in the heritability of *EduYears*.

### 3.6 Cohort Characteristics and Genetic Correlation of *EduYears*

#### Empirical methods

To explore which cohort characteristics are associated with the genetic correlation of *EduYears*, we estimate a weighted linear regression model with inverse-variance weighting. The dependent variable is the estimated genetic correlation for a cohort pair (see Section 3.3). There are 933 observations included in the model. We include four continuous independent variables, each of which equals the absolute value of the difference in the corresponding cohort characteristic between the two cohorts. As we did earlier when calculating the mean genetic correlation, we use the node-jackknife variance estimator<sup>39</sup> to calculate standard errors, since genetic correlation may be correlated across all observations that share a cohort in common.

---

the exploratory cohort-level analyses described in Sections 3.5 and 3.6. However, when we re-did the analysis using our larger data set, we learned that the results are highly sensitive to the cut-points used to define the dichotomous variables. We prefer the cohort-level approach because it avoids dividing cohorts at arbitrary cut-points. However, we caution readers that the empirical approach we take in Sections 3.5 and 3.6 was not the one we intended to conduct with our larger data set.

## Results

**Supplementary Table 25**, Panel B, shows results from this exploratory analysis. The estimated intercept from this regression is 0.868 (SE = 0.228), which means that 0.868 is the expected genetic correlation of *EduYears* between cohorts that have identical values (absolute value of the difference = 0) on all four measures of cohort characteristics. This intercept is statistically distinguishable from zero ( $P$  value =  $1.54 \times 10^{-4}$ ), but not from one ( $P$  value = 0.563).

The expected genetic correlation of *EduYears* between cohorts declines as the difference in mean birth years of the cohorts grows ( $\beta = -0.009$ ; SE = 0.004;  $P$  value = 0.037). That is, the genetic correlation of cohorts with similar mean birth years, such as 23andMe (mean birth year = 1961) and Geisinger (1960) is expected to be higher than that between, for example, Add Health (1979) and either 23andMe or Geisinger. Indeed, the estimated genetic correlation of *EduYears* for 23andMe and Geisinger is 0.929 (SE = 0.131), while it is lower between Add Health and 23andMe (0.751, SE = 0.089) and between Add Health and Geisinger (0.787, SE = 0.214).

The other three variables in the regression of genetic correlation do not play statistically significant roles in cross-cohort genetic correlation of *EduYears* in our data (difference in number of response options for *EduYears*:  $\beta = -0.001$ , SE = 0.006,  $P$  value = 0.935; difference in average education in the country:  $\beta = 0.016$ , SE = 0.042,  $P$  value = 0.702; difference in top income shares in the country:  $\beta = -0.012$ , SE = 0.014,  $P$  value = 0.368). As indicated by the  $R^2$ , together, the four independent variables explain 19.0% of variation in the genetic correlation estimates across cohorts. As with the analysis of heritability, since much of the variation in the estimates of genetic correlation is due to sampling variance, this should be thought of as a lower bound of the variance of the *true* genetic correlation explained by our covariates.<sup>j</sup>

---

<sup>j</sup> Following the suggestion of a referee, we also conducted a post hoc analysis in which we assessed whether genetic correlation differs depending on whether the cohorts are from the same geographic region or not. We partitioned the cohorts into the following geographic regions: Australasia, Baltic countries, Nordic countries, North America, Southern Europe, United Kingdom, or Western Europe. Of the  $n = 933$  genetic correlations estimated,  $n = 175$  (19%) are from cohorts from the same region, while  $n = 758$  (81%) are from cohorts from different regions. We then followed a similar analytic strategy to that used in Supplementary Section 3.6. We regressed pairwise genetic correlations on the four main cohort characteristics as well as a dichotomous variable coded 0 if the cohorts are from unmatched regions and 1 if they are from the same region, we applied inverse-variance weighting, and we estimated standard errors and  $P$  values using the node-jackknife estimator<sup>39</sup>. The coefficient on the indicator for matched regions

### 3.7 UK Biobank Analyses

In our final set of analyses, we again assess the relationships between cohort characteristics and the heritability and genetic correlation of *EduYears*. Unlike the analysis above, however, here we analyze individual-level data from a single cohort, the full release of the UK Biobank, which contains data from over 400,000 respondents. This analysis holds constant potentially confounding country-level variables that we were not able to control for in the cross-cohort analyses described above. The cohort characteristics we are able to explore within the UK Biobank are, however, more limited than those assessed in the cross-cohort analysis, since the characteristics studied must vary within the UK. Thus, in this section we assess the associations of the heritability and genetic correlation of *EduYears* with two of the four characteristics we explore above: the number of response options for *EduYears* and birth year.

#### Empirical methods

Our analytic strategy proceeds in two steps. First, we compare the heritability and genetic correlation of *EduYears* estimated using the full set of response options provided by the UK Biobank to that obtained when simulating a scenario in which fewer options are provided. Second, we divide the UK Biobank cohort in half in order to estimate the heritability of *EduYears* and the genetic correlation of *EduYears* between two subsamples: one comprised of respondents born in the first half of the twentieth century, and the other comprised of those born in later years. These methods are described in greater detail below. We note that although the UK Biobank is one of the cohorts included in the cross-cohort analysis above, the analysis here provides an independent source of evidence because we are exploiting individual-level variation within the cohort.

*Number of response options for EduYears:* The UK Biobank’s question about educational attainment includes seven categories for respondents to choose from. For our “original” variable, these categories are mapped onto years of education using the ISCED scale as follows: none of the above (no qualifications) = 7 years of education; CSEs or equivalent = 10 years; O levels/GCSEs or equivalent = 10 years; A levels/AS levels or equivalent = 13 years; other professional qualification = 15 years; NVQ or HNC or equivalent = 19 years; college or university degree = 20 years of education.

The “coarsened” measure of years of education simulates a scenario in which only three response options were provided. The first option is to report less than a high school education (by UK standards, less than A or AS levels). For this option, we combine those

---

is noisily estimated and not statistically distinguishable from 0 ( $\beta = -0.094$ ,  $SE = 0.100$ ,  $P$  value = 0.347); other results are substantively unchanged.

who reported no qualifications, O levels/GSCEs or equivalent, and CSEs or equivalent, and code them as having attained 10 years of education. Second, respondents could report the equivalent of a high school degree (A levels/AS levels or equivalent), which we coded as 13 years of education. Finally, respondents could report any post-secondary school training. For this category, we combine those with professional qualifications, NVQ, HNC or equivalent, or a college or university degree, and code them as having attained 19 years of education. The categories contain 152,690, 24,214, and 265,279 individuals, respectively.

We perform GWA studies of both the original and coarsened years of education variables, both of which include data from 442,183 respondents. We then calculate the heritabilities of and the genetic correlation between the original and coarsened phenotypes using LD Score regression<sup>8,14</sup>.

*Birth years:* As outlined above, we divide the UK Biobank at the median birth year to form two subsamples. The early birth year subsample includes the 230,362 respondents born between 1934 and 1950 (mean birth year = 1945), while the later birth year subsample includes 211,821 respondents born between 1951 and 1970 (mean birth year = 1958). Within each subsample, we run GWA studies of years of education. Using LD Score regression<sup>8,14</sup>, we then calculate the heritability of *EduYears* in each subsample and the genetic correlation of EA between the subsamples.

For comparison, we run parallel analyses of genetic correlation for two additional phenotypes: height and body mass index (BMI). Recall that we hypothesized that an imperfect genetic correlation of *EduYears* between cohorts with differing mean birth years would be driven by societal changes in the motivations and requirements for schooling. We consider height a negative control because we expect recent societal shifts to have had a much smaller effect on the genetic factors that contribute to height. Indeed, we expect a near-perfect genetic correlation of height between the two subsamples. We expect the genetic correlation of BMI between the subsamples to fall somewhere between that for *EduYears* and that for height. BMI is influenced in part by biological factors that are not affected by societal change, but shifts over time in access to different types of food and in knowledge about the dangers of high BMI might lead to imperfect genetic correlation across earlier and later birth cohorts. With data available for height, there are 229,845 individuals with below-median birth year and 211,429 individuals at or above the median birth year. For BMI, the corresponding sample sizes are 229,580 and 211,246 individuals.

## Results

*Number of response options for EduYears:* Results regarding number of response options mirror those of the cross-cohort analysis described above. We estimate heritability of 0.137 (SE = 0.004) for the original variable (as indicated in **Supplementary Table 22**) and 0.106 (SE = 0.004) for the coarsened variable. We strongly reject the null hypothesis of equal heritability between these values ( $P$  value =  $3.47 \times 10^{-9}$ ).<sup>k</sup> Thus, the estimated heritability of *EduYears* is lower when fewer response options are provided. The genetic correlation between *EduYears* as recorded in its original form and *EduYears* in its coarsened form is nearly perfect, at 0.996 (SE = 0.001).

*Birth years:* The estimated heritability of *EduYears* among those born in earlier years is 0.157 (SE = 0.006) while that for later years is 0.134 (SE = 0.006). This is a statistically significant difference ( $P$  value = 0.003). Note that in the cross-cohort analysis, we found no evidence of a statistically significant difference in heritability across birth cohorts.

The genetic correlation of *EduYears* between the earlier and later birth year subsamples is nearly perfect, at 0.997 (SE = 0.018). This result does not support findings from the cross-cohort analysis. There, we found that the estimated genetic correlation of *EduYears* may be significantly lower for pairs of cohorts with different average birth years.

Between subsamples within the UK Biobank, we also find nearly perfect genetic correlations of BMI (0.982, SE = 0.010) and height (0.989; SE = 0.008). Thus, relative genetic effects estimated in these two subsamples of the UK Biobank appear approximately equal for all three phenotypes.

### 3.8 Concluding Discussion

In this section, we studied variation in the heritability and genetic correlation of *EduYears* across cohorts. This analysis was motivated by empirical evidence of heterogeneous effect sizes for the lead SNPs, a result that could be driven by differences in heritability or imperfect genetic correlation across cohorts. One reason that such heterogeneity may be of interest is that the predictive power of a polygenic score (PGS) depends on these quantities, as shown in a derivation by de Vlaming et al.<sup>36</sup> Specifically, the expected predictive power of a PGS declines with the heritability of the phenotype in

---

<sup>k</sup> We test for equality using a standard two-sample  $t$ -test to compare differences in means ( $t = 5.908$  with 442,182 degrees of freedom). This test is conservative because it assumes that the two samples are independent. In fact however, since the two heritability estimates are based on the same sample, the two estimates will be positively correlated, and therefore the  $P$  value we calculate is an upper bound on the true  $P$  value.

the prediction cohort, and it is lower when the genetic correlation of the phenotype between the discovery and prediction cohorts is imperfect.

Across the 48 cohorts included in our analysis, we estimate a mean heritability of 0.147 (SE = 0.009) and a standard deviation of 0.023 (SE = 0.048). As indicated by the large standard error on the standard deviation point estimate, we are not well-powered to detect heterogeneity of SNP heritability across cohorts. Nonetheless, we view our approach as a promising way of obtaining evidence that is complementary with earlier twin-study-based evidence of heterogeneity in the broad-sense heritability of educational attainment across cohorts<sup>25</sup>. In our exploratory analyses of the association of cohort characteristics with the heritability of *EduYears*, the only robust predictor of estimated heritability is the precision of the survey measure of *EduYears*. Specifically, in both the cross-cohort analysis and the within-UKB analysis, when a cohort has a measure of *EduYears* with more response options, the estimated heritability of *EduYears* tends to be larger in that cohort. This finding is consistent with finer measures being less prone to measurement error.

Due to gene-by-environment interactions, the heritability of *EduYears* could also differ depending on the institutional environments faced by cohort respondents. We found some evidence of this in the cross-cohort analysis, as income inequality (measured with top income shares) was associated with increasing heritability. We are unable to verify this finding in the UK Biobank analysis, however. Also, in the UK Biobank analysis, we find that the heritability of *EduYears* is lower in the more recent birth cohort, a result that is not observed in the cross-cohort regression.

To our knowledge, this is the first study to systematically investigate not only variation in SNP heritability of educational attainment, but also to systematically study variation in the genetic correlation of educational attainment across pairs of cohorts. We find a mean genetic correlation of 0.723 (SE = 0.124), which is substantially smaller than one ( $P$  value = 0.026). This finding indicates that the average pair of cohorts considered in this analysis does not have a perfect genetic correlation of *EduYears*, and provides further evidence of heterogeneity in genetic effects on *EduYears* across cohorts.

As with heritability, genetic correlation across pairs of cohorts might depend on the institutional environments faced by the cohorts considered. In particular, the motivations and requirements to advance through the educational system may have shifted across the twentieth century. As a result, the genetic correlation of *EduYears* may be imperfect across pairs of cohorts with respondents born in different time periods. The cross-cohort analysis suggests precisely this, as pairs of cohorts with very different average birth years

evinced lower genetic correlations of *EduYears* than cohort pairs with more similar average years of birth.

However, this result is not supported by the UK Biobank analysis, where we find a nearly perfect genetic correlation of *EduYears* across respondents born in the first and second halves of the twentieth century. This suggests that birth year may be a proxy for some characteristic of the cohort that is related to genetic correlation. As an illustrative example, from **Supplementary Table 24**, Panel B, we see that the average birth year and average education in the country variables have a correlation of 0.61. Suppose that differences in mean educational attainment explains some of the attenuation in cross-cohort genetic correlation, but that differences in birth year between cohorts are a better proxy for the actual differences in mean cohort educational attainment than the measure of educational attainment that we used in this analysis. Under this scenario, we would observe the sort of patterns presented above.

Despite differences in the specific findings, both the cross-cohort and UK Biobank analyses suggest that the predictive power of a PGS for *EduYears* may vary across cohorts that differ in average birth year. Specifically, the UK Biobank results imply a lower predictive power in later-born cohorts because the estimated heritability of *EduYears* is lower for UK Biobank respondents born in the second half of the twentieth century, compared to those born in earlier years. The cross-cohort results imply a lower predictive power in cohorts with mean birth years that are dissimilar to that of the meta-analysis from which the PGSs are derived. Both results may therefore be consistent with those of Okbay et al.<sup>1</sup>, which finds that the predictive power of a PGS for *EduYears* was higher among earlier-born respondents to the Swedish Twin Registry than among later-born respondents. It may be that the heritability of *EduYears* is in fact lower among the later-born respondents. It could also be the case that the genetic correlation of *EduYears* with the meta-analyzed sample is lower for the later-born respondents, as that meta-analysis relied heavily on cohorts with average years of birth in the *early* twentieth century.<sup>1</sup>

Of course, cohort characteristics not studied here (or better measures of the characteristics we study) might explain more of the variation in the heritability and genetic correlation of *EduYears*. Further identifying and exploring the characteristics that

---

<sup>1</sup> Our findings are also consistent with those of (350)<sup>356</sup>, which finds that the estimated heritability of educational attainment is 40% lower assuming homogeneous genetic effects across cohorts than when allowing effects to vary by birth year and country. After conducting simulations that rule out alternative explanations, Tropf et al. argues that differences in genetic effects result in imperfect genetic correlation across environments, suppressing heritability estimates in pooled data.



explain the heterogeneity in genetic effects for educational attainment is an important area for future research.

## 4. X-Chromosome Analysis

### 4.1 Introduction

Our primary meta-analysis of *EduYears* was restricted to autosomal SNPs. In this section, we describe the results from supplemental association analyses of SNPs on the X chromosome that we performed in our two largest cohorts, UK Biobank and 23andMe. The X chromosome is understudied in genome-wide association studies of complex traits. In addition to gene and gene-variant discovery on the X chromosome, we were interested in addressing a number of specific questions: (i) what is the total contribution of common SNPs on the X chromosome to phenotypic variation in *EduYears*, (ii) what is the relative contribution of common SNPs to phenotypic variation in *EduYears* in males and females, and (iii) what is the genetic correlation of male and female *EduYears* due to common SNPs on the X chromosome? (iv) how does the contribution of common SNPs on the X chromosome compare to the contribution of common SNPs of autosomes similar in length, or similar in terms of the effective number of loci.

We begin by establishing some notation and a framework that will be helpful for motivating and interpreting subsequent analyses. Next, we describe quality control and imputation of the UK Biobank data and association analyses (both mixed-sex and single-sex) we subsequently conducted in this sample. We use the results from the sex-stratified association analyses to estimate the amount of dosage compensation and the male-female genetic correlation. Finally, we report the results from a meta-analysis of summary statistics from mixed-sex association analysis conducted in 23andMe and UK Biobank (using identical allele coding across the two datasets).

### 4.1 Notation and Theoretical Framework

*Notation.* Let  $y_m$  and  $y_f$  denote the phenotype value for a male and a female, respectively, and let  $x_m \in \{0,1\}$  and  $x_f \in \{0,1,2\}$  denote the allele counts in males and females. Let  $\beta_f$  denote the coefficient from the population regression of  $y_f$  on  $x_f$ , with  $b_f$  its sample analog, and define  $\beta_m$  and  $b_m$  similarly for males. Finally, denote the minor allele frequency by  $p$ . We have that  $\text{Var}(x_m) = p(1-p)$  and, under Hardy-Weinberg assumptions,  $\text{Var}(x_f) = 2p(1-p)$ .

*Dosage Compensation.* The genetic variance contributed by a SNP on the X chromosome in females is  $\text{Var}(\beta_f x_f) = \beta_f^2 \text{Var}(x_f) = 2p(1-p)\beta_f^2$ , and that in males is  $\text{Var}(\beta_m x_m) = \beta_m^2 \text{Var}(x_m) = p(1-p)\beta_m^2$ . The relationship between the effect sizes  $\beta_f$  and  $\beta_m$ , and between the genetic variances contributed in males and females, depends on

the amount of sex-linked dosage compensation (X-inactivation) in females<sup>40</sup>. Dosage compensation can be parameterized as  $\beta_m = d\beta_f$ , where  $1 \leq d \leq 2$ . In the absence of dosage compensation ( $d = 1$ ),  $\beta_m = \beta_f$ , and the variance contributed by the SNP in males is  $\text{Var}(\beta_m x_m) = \beta_m^2 \text{Var}(x_m) = p(1-p)\beta_f^2$ , which is half that in females. Under full dosage compensation ( $d = 2$ ),  $\beta_m = 2\beta_f$ , and the variance contributed by the SNP in males is  $\text{Var}(\beta_m x_m) = \beta_m^2 \text{Var}(x_m) = 4p(1-p)\beta_f^2$ , hence twice that in females. Finally, under partial dosage compensation, the variance contributed in males is  $d^2 p(1-p)\beta_f^2$ .

*Sex-Stratified Association Analyses in UK Biobank.* As described in detail in the next subsection, we conducted sex-stratified association analyses in UK Biobank. We use the results to conduct two analyses that are informative about the amount of dosage compensation. First, under the maintained assumption that  $\beta_m = d\beta_f$  across all X-chromosome SNPs with a common value of  $d$ , we can estimate  $d$  using the association statistics. The expected mean  $\chi^2$  statistic on the X-chromosome is:

$$\mathbb{E}[\chi_i^2] = 1 + \frac{N_i h_i^2}{M_{\text{eff}}}, \quad (4.1)$$

where  $i \in \{m, f\}$  indicates males or females,  $h_i^2$  is the SNP heritability for the X chromosome,  $N_i$  is the GWAS sample size, and  $M_{\text{eff}}$  is the effective number of loci (which is assumed to be the same in males and females). Rearranging to solve for  $\gamma \equiv \frac{h_m^2}{h_f^2}$  and replacing the population moments with their sample analogs yields an estimator for the ratio of male-to-female heritability:

$$\hat{\gamma} = \frac{(\hat{\chi}_m^2 - 1)N_f}{(\hat{\chi}_f^2 - 1)N_m}. \quad (4.2)$$

In what follows, we refer to  $\gamma$  as the dosage compensation ratio. The ratio takes on a value between 0.5 (zero dosage compensation) and 2 (full dosage compensation). Second, we use the results from the sex-stratified analyses to test a key implication of the assumption that  $\beta_m = d\beta_f$ , namely that the male-female genetic correlation on the X chromosome is unity. The expectation of the product of the  $z$ -statistics from the female and male analyses is

$$\mathbb{E}[z_m z_f] = \frac{r_g h_m h_f (\sqrt{N_m N_f})}{M}, \quad (4.3)$$

where  $r_g$  is the male-female genetic correlation. Using equation (4.1) we have  $h_i = \sqrt{\frac{(\chi_i^2 - 1)M}{N_i}}$  for  $i = m, f$ . Substituting into equation (4.3), rearranging, and replacing population moments by their sample analogs, we obtain an estimator for the genetic correlation:

$$\hat{r}_g = \frac{\widehat{\bar{z}_m \bar{z}_f}}{\sqrt{(\hat{\chi}_f^2 - 1)(\hat{\chi}_m^2 - 1)}}. \quad (4.4)$$

Standard errors are calculated using a block jackknife procedure as follows. We first defined  $B = 1,000$  blocks of contiguous SNPs across the X chromosome. For each block  $k$ , we calculate an estimate  $r_g^{(k)}$  of the genetic correlation as in equation (4.4) using all SNPs except for those included in the  $k$ -th block. The standard error is then calculated with following formula:

$$SE(\hat{r}_g) = \sqrt{\frac{B-1}{B} \sum_{k=1}^B (\hat{r}_g - r_g^{(k)})^2}. \quad (4.5)$$

Using equations (4.4) and (4.5), we calculate a  $z$ -statistic and test the null hypothesis that  $r_g = 1$ .

*Joint Analyses in UK Biobank.* We also performed two joint analyses in pooled samples of males and females in UK Biobank. In both, female allele counts were coded 0/1/2. In our first joint analysis, which we will sometimes refer to below as the full-dosage-compensation analysis, males were coded 0/2. In the second, which we refer to as our zero-dosage-compensation analysis, males were coded 0/1. To interpret the results of these analyses (and justify their names), it is useful to examine their statistical properties under alternative assumptions about the true value of the dosage-compensation parameter.

In a joint analysis of males and females, the phenotype is regressed on a genotype variable equal to  $x_f \in \{0,1,2\}$  for females and  $cx_m$  for males, with  $c = 1$  in the zero-dosage-compensation analysis or  $c = 2$  in the full-dosage-compensation analysis. Under the simplifying assumption that the residual variance is exactly one in both males and females, the coefficient from the pooled analysis can be written as a weighted sum of the regression coefficients  $b_f$  and  $b_m$  that would have been estimated in separate analyses:

$$b_{\text{joint}} = \frac{\frac{cb_m}{\text{Var}(b_m)} + \frac{b_f}{\text{Var}(b_f)}}{\frac{c^2}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}}. \quad (4.6)$$

Under the assumption that  $\beta_m = d\beta_f$ , the expectation of the estimator is:

$$E(b_{\text{joint}}) = \frac{[\frac{cd}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}]}{\frac{c^2}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}} \beta_f. \quad (4.7)$$

Thus, the estimator is unbiased when  $c = d$ . Given that we set  $c = 1$  or  $c = 2$  in our UK Biobank association analyses, the  $c = 1$  analysis is only unbiased under zero dosage compensation, and the  $c = 2$  analysis is only unbiased under full dosage compensation. Moreover, it can be shown that an optimally weighted meta-analysis of association results from sex-stratified analyses will weight the sex-specific estimates as in equation (4.6), setting  $c = d$ . That is, the joint analysis with  $c = d$  is optimal in the sense that the resulting estimator has the lowest variance among the class of unbiased estimators.

Because we will use it later, we derive here the noncentrality parameter (NCP) for the  $\chi^2$  test of the null hypothesis  $H_0: \beta_m = \beta_f = 0$  against the alternative  $H_1: \beta_m = c\beta_f \neq$

0. Recall that this *NCP* is defined as  $\left( \frac{b_{\text{joint}|H_1} - b_{\text{joint}|H_0}}{\sqrt{\text{Var}(b_{\text{joint}})}} \right)^2$ , and note that  $\text{Var}(b_{\text{joint}}) =$

$\frac{1}{\frac{c^2}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}}$ . Hence,

$$NCP = \frac{[\frac{cd}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}]^2}{\frac{c^2}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}} \beta_f^2. \quad (4.8)$$

At the optimal estimator ( $c = d$ ), the NCP is

$$NCP = [\frac{d^2}{\text{Var}(b_m)} + \frac{1}{\text{Var}(b_f)}] \beta_f^2. \quad (4.9)$$

We note that these derivations imply that summary statistics from X-chromosome analysis performed in males and females separately can be meta-analysed using weights that give results that are almost identical to those that would result from a joint male-female analysis on individual-level data using SNP coding that corresponds to any

assumed model of dosage compensation (we verified this prediction empirically using two alternative coding schemes, one corresponding to full dosage compensation and one corresponding to no dosage compensation).

## 4.2 UK Biobank: Imputation, Quality Control and Association Analyses

*Selection of UK Biobank European Sample.* Ancestry assignment was performed using the autosomal markers. UK Biobank-provided genotype probabilities were used to hard-call the genotypes. If the genotype posterior probability was  $\leq 0.9$ , then the genotype was treated as missing. We then filtered the variants to retain only those which were hard-called in  $\geq 95\%$  of individuals. All subsequent analyses were further restricted to HapMap3 SNPs with missingness below 5% and minor allele frequency above 1%.

To identify a sample of conventionally unrelated European-ancestry respondents, we calculated the first two principal components (PCs) from the 2,504 participants with known ancestries from the 1000 Genomes Project. We then projected UK Biobank respondents onto those PCs, obtaining the loadings of each UK Biobank respondent on each PC. We then assigned each individual to one of five super-populations in the 1000 Genomes data: European, African, East Asian, South Asian and Admixed. Our algorithm for doing so calculated, for each respondent, the probability of membership to the European super-population conditional on their PC coordinates. The 456,426 out of the original 487,409 respondents who had a probability of membership  $> 0.9$  for European were assigned to the European super-population.

Next, to obtain an estimation sample of conventionally unrelated individuals, we estimated a genetic relatedness matrix (GRM) for individuals in the subsample of Europeans. We iteratively dropped one member from each pair of individuals whose estimated relatedness exceeded 0.05, until no two individuals with a relatedness above 0.05 remained in the sample. This restriction resulted in our sample of 348,580 conventionally unrelated Europeans.

In this sample, we used the software flashPCA<sup>41</sup> to calculate PCs that we use as controls in our association analyses. While we calculated our own PCs instead of using the UK Biobank-provided PCs, we calculated them from the same set of genotyped autosomal SNPs that the UK Biobank used for calculating their PCs. The SNPs in this set have been pruned with LD threshold  $r^2 = 0.1$  and have had high LD regions removed (as described in the supplementary material S3.3.1 and S.3.3.2 and Table S12 of Bycroft et al.<sup>42</sup>). To this set of SNPs, we only retained SNPs satisfying each of the following quality control criteria: missingness  $< 5\%$ , MAF  $> 1\%$ , and HWE  $P$  value  $> 10^{-6}$ . These steps left us with 137,102 SNPs for the PC calculations.

*Imputation.* Imputed genotypes for the X chromosome were not included in the data officially released by UK Biobank. We therefore imputed the data ourselves. We imputed 3,351,438 variants in the non-pseudo-autosomal region (PAR) and 100,087 markers in the PAR1 region (from the start of the X chromosome to bp 2,699,507) using the 1000 Genomes Project as our reference panel. Prior to imputation, we selected SNPs used as X-chromosome phasing input by UK Biobank and applied a few additional quality-control filters. We dropped respondents: (i) classified as heterozygosity or missingness outliers, (ii) whose reported sex did not match their biological sex, (iii) with putative sex chromosome aneuploidy (i.e., abnormal number of sex chromosomes), and (iv) excluded from kinship inference. We then identified the set of non-PAR SNPs that satisfied the following restrictions: (i) HWE  $P$  value  $> 10^{-6}$  in the subsample of females, (ii) MAF  $> 10^{-4}$  in males and females, and (iii) SNP call rate  $> 95\%$  in males and females. These filters left a set of 15,424 common variants that we used for genotype imputation in 455,617 European-ancestry subjects. We also imputed some markers from the PAR1 region, but all analyses below are based on SNPs from the non-PAR region.

*Quality Control of Imputed Markers.* We performed quality-control analyses of a set of 1,059,233 SNPs from the non-PAR region with imputation info score  $> 0.3$  and MAF  $> 0.0001$ . In this sample, we found substantial differences in allele frequencies of imputed variants between males and females that were not present among the originally genotyped SNPs. We subsequently excluded markers with HWE  $P$  value  $< 10^{-6}$  in females, leaving a set of 1,024,430 markers. Among these markers, the allele frequencies were much more consistently aligned across males and females. In analyses not shown here, we found that the consistency does not improve noticeably with even tighter HWE  $P$  value or info thresholds.

*Estimation Sample.* In all UK Biobank analyses, our *EduYears* phenotype is defined exactly as in the autosomal analyses<sup>1</sup>. Before our final association analyses, we dropped from our sample of 348,580 conventionally unrelated Europeans (whose construction was described above) individuals who failed the inclusion criteria for imputation (sex/gender mismatch, heterozygosity missingness outlier, or putative sex chromosome aneuploidy). We also dropped participants who were not born in the UK or who had withdrawn their consent for their data to be used. Applying all these exclusions leaves us with our final sample of  $N = 329,358$  conventionally unrelated respondents of European-ancestry. In our sex-stratified association analyses, we residualize our phenotype on 10 PCs, indicator variables for year of birth, and indicator variables for genotype-measurement batch. In our joint association analyses described below, we additionally residualize on interactions between these covariates and sex, as well as a main effect for sex.

### 4.3 UK Biobank Association Results

In our sex-stratified analyses, we split our estimation sample into females only ( $N = 176,750$ ) and males only ( $N = 152,608$ ) but otherwise used identical methods; males were coded as 0/1 as females as 0/1/2 (i.e., the SNP effects were estimated in units of per allele). We also conducted two joint male-female analyses. In both joint analyses, female allele counts were coded 0/1/2. In the first analysis (“Full DC”), we coded males 0/2. In the second (“No DC”), we coded males 0/1. Our analyses were based on 1,024,430 SNPs in the non-PAR region. **Supplementary Table 26** provides a summary overview of the results from the Males, Females, Full DC, and No DC analyses. For all SNPs with allele frequency above 1%, 0.1% or 0.01%, and each of the four association analyses, the table separately reports the sample size, the mean test statistic, and the maximum test statistic.

*Sex-Stratified Analyses.* The sex-stratified analyses yielded similar findings overall across males and females. Using equation (4.2), it is straightforward to calculate the implied dosage compensation ratio from the sample sizes and mean test statistics reported in **Supplementary Table 26**. For SNPs with  $MAF > 1\%$ , our estimate is 1.05, but estimates based on other MAF cutoffs are very similar. Equation (4.2) shows that the estimate of dosage compensation does not depend on the parameter  $M_{\text{eff}}$  because the terms in the numerator and denominator cancel. However,  $M_{\text{eff}}$  is needed calculate the standard error of  $\hat{\gamma}$ .<sup>m</sup> We estimated  $M_{\text{eff}}$  by calculating the reciprocal of the variance of the off-diagonal elements of a genetic relatedness matrix based on SNPs on the X chromosome (see Goddard et al.<sup>43</sup> for a theoretical derivation and discussion of identifying assumptions). We calculated the GRM for SNPs with  $MAF > 1\%$  and  $HWE P > 10^{-6}$  in a sample of approximately unrelated females. This gave an estimate of  $M_{\text{eff}}$  of  $\sim 1,300$ , which we take to be the effective number of independent markers on the X chromosome in all that follows. For this value of  $M_{\text{eff}}$ , the estimated standard error of  $\hat{\gamma}$  is approximately 0.15, implying a 95% confidence interval ranging from 0.76 to 1.33.

Next, we used equation (4.4) to estimate the genetic correlation between males and females. Reassuringly, the estimate is close to unity irrespective of the MAF cutoff used

---

<sup>m</sup> Assuming that the individual test statistics are distributed as a non-central  $\chi^2$  with expected value given by equation (4.1) and variance  $2[1 + h_i^2 N_i / M_{\text{eff}}]$ , the variance of the mean test statistic across the chromosome is approximately  $(2/M_{\text{eff}})[1 + 2(\hat{\chi}^2 - 1)]$ , and the variance of the dosage compensation ratio is approximately  $\gamma^2 \frac{N_f^2}{N_m^2} \left[ \frac{\text{Var}(\hat{\chi}_m^2)}{(\hat{\chi}_m^2 - 1)^2} + \frac{(\hat{\chi}_m^2 - \hat{\chi}_f^2)^2 \text{Var}(\hat{\chi}_f^2)}{(\hat{\chi}_m^2 - 1)^2 (\hat{\chi}_f^2 - 1)^2} \right]$ .



to determine SNP inclusion. Limiting the analysis to SNPs with MAF above 1%, the genetic correlation estimate is  $r_g = 1.01$  (jackknife SE = 0.05).

*Results from Joint Analyses.* **Supplementary Table 26** reports some summary statistics from our two joint analyses. Results from the first, which is optimal under zero dosage compensation, are shown in the row labelled Zero Dosage Compensation (0-1). Results from the second analysis, which is optimal under full dosage, are shown in the row labelled Full Dosage Compensation. The Zero DC analysis yielded three approximately independent lead SNPs at genome-wide significance, whereas the Full DC analysis yielded four. The mean inflation statistics are marginally higher for the Full Dosage Compensation analysis at all MAF cutoffs, but the differences are small. Evaluating equation (4.8) at  $d = 1.45$  (SE = 0.12), we find that the noncentrality parameter of the Zero Dosage Compensation model is marginally lower at  $c = 1$  (corresponding to the Zero DC analysis) than at  $c = 2$  (corresponding to the Full DC analysis). Both joint analyses were conducted using a value of  $c$  that is quite far away from 1.45, the value of the dosage-compensation parameter that we estimate would maximize statistical power. However, we show below that the actual loss of efficiency due to the inefficient weighting is small.

#### 4.4 Association Analysis in 23andMe

We obtained summary statistics from association analyses of SNPs on the X chromosomes conducted among research participants from 23andMe ( $N = 365,536$  individuals). These analyses were conducted from a joint male-female analysis with male genotypes coded 0/2. With the exception of allele coding, all other major aspects of the analyses were identical to those described for the autosomal analyses in **Supplementary Section 1**; see **Supplementary Table 17** for details on the phenotype, **19** for information about the association model used, and **18** for information about imputation.

#### 4.5 Quality Control of UK Biobank and 23andMe Results

Both sets of association results underwent a set of quality-control filters similar to those described in **Supplementary Section 1**. Association results for 239,654 SNPs in 23andMe and 284,939 SNPs in UK Biobank remained after application of these filters. We subsequently dropped a small number of SNPs with male-female allele frequency differences above 0.005 in UK Biobank. These restrictions leave 238,249 SNPs for meta-analysis in 23andMe and 284,068 SNPs in UK Biobank.

#### 4.6 Meta-Analysis of UK Biobank and 23andMe Results ( $N = 694,894$ )

The X-chromosome results from the UK Biobank and 23andMe were meta-analyzed using sample-size weighting in METAL<sup>5</sup>. We used association results from the UK Biobank obtained with the same coding of males (0/2) as that used by 23andMe. Only SNPs that were present in both results files were used. This final restriction leaves 205,865 SNPs for meta-analysis in a combined sample of  $N = 694,894$  individuals. To adjust the test statistics for bias, we inflated the standard errors by the LD Score regression intercept from our main autosomal analysis ( $\sqrt{1.113}$ ), and we calculated  $P$  values using these adjusted test statistics.<sup>n</sup> Applying our clumping algorithm, we found 10 approximately independent SNPs at genome-wide significance. The mean  $\chi^2$  test statistic, calculated before (after) inflation adjustment, was 2.72 (2.45). For SNPs with minor allele frequency above 1%, the analogous number is 2.86 (2.57). **Supplementary Figure 5** shows Manhattan and quantile-quantile plots from the meta-analysis.

The 10 lead SNPs we find on the X chromosome are listed in **Supplementary Table 4**, along with the gene closest to each SNP. Two of the SNPs are closest to the same gene, *FAM47A*, so there are 9 unique genes listed in the table. It is noteworthy that 2 out of the 9, *HUWE1* and *GPC3*, are believed to be genes in which *de novo* mutations cause intellectual disability<sup>44</sup>. Although it would be of interest to test more systematically for evidence that the lead SNPs we identify on the X chromosome are enriched for intellectual-disability-related genes relative to the lead SNPs we identify on the autosomes, we do not pursue such an analysis because the number of genes identified on the X chromosome is small.

#### 4.7 Comparison to Autosomes

To allow comparisons of the results from our X chromosome meta-analysis to those from the autosomal meta-analysis, we began by running a new autosomal meta-analysis restricted to 23andMe and UK Biobank. Because the autosomal GWAS in UK Biobank was not limited to conventionally unrelated individuals, the resulting autosomal meta-analysis was based on ~808,000 individuals, an approximately 16% larger sample size than the one used in our X chromosome meta-analysis. As we describe below, all comparisons are based on test statistics that are adjusted for the difference in sample size.

We compared the autosomal and X-chromosomal meta-analyses along several dimensions, the first of which is the number of lead SNPs on each chromosome. To adjust for sample-size differences, we first inflated all standard errors in the autosomal

---

<sup>n</sup> As a robustness check, we also examined the results when we inflate the standard errors by the LD Score regression intercept ( $\sqrt{1.099}$ ) from an autosomal meta-analysis restricted to UKB and 23andMe, a sample more comparable to our X chromosome meta-analysis sample. In that case, there is one additional lead SNP (11 rather than 10): rs5951458.

meta-analysis and calculated adjusted test statistics based on these inflated test statistics. Since standard errors decline at a rate that is inversely proportional to the square root of sample size, we used the adjustment factor of  $\sqrt{808,000/694,894} \approx 1.079$ . We subsequently applied our clumping algorithm to these results and calculated the number of lead SNPs per chromosome. For the purposes of this comparison, the number of lead SNPs on all chromosomes (X and autosomes) were always calculated without adjusting test statistics for inflation due to stratification biases or cryptic relatedness.

The results are shown in **Supplementary Table 5**. The SNPs in our analysis of the X chromosome span a total distance of 152 MegaBase Pairs (Mbps), and by this metric, it is most similar to chromosomes 6 through 10 (mean 150, range 135 to 171). Even adjusting for sample-size differences, the number of lead SNPs on the X-chromosomal analysis is about one third as large as the number of SNPs on autosomes of similar length: we identify 12 lead SNPs on the X chromosome, whereas on chromosomes 6 through 10, the average number of lead SNPs is 33 (range 27 to 41).

One possible factor contributing to the discrepancy is that our meta-analysis of SNPs on the X chromosome likely used suboptimal weights. Using equations (4.8) and (4.9), we calculated that the power of the meta-analysis we conducted in our sample of  $N = 694,894$  was equal to an optimally weighted meta-analysis conducted in a sample of  $N = 709,964$ . The fact that this difference is small implies that the sub-optimality of our weighting scheme contributes minimally to the observed discrepancy.

A second potentially contributing factor is that the effective number of independent markers on the X chromosomes may be low compared to the effective number of independent markers on autosomes of similar length. On the one hand, this would boost the signal to detect associated loci by inflating the LD Score of SNPs on the X chromosome; at the same time, due to the large available sample and the high polygenicity of *EduYears*, a lower effective number of markers will mean that each lead SNP may tag a larger number of independently associated SNPs, leading to fewer lead SNPs being identified by the clumping algorithm. To test which of these effects dominates, we estimated the effective number of loci,  $M_{\text{eff}}$ , for each chromosome, using methods identical to those described above for the X chromosome, and compared the number of lead SNPs to the chromosome-specific SNP heritability for each chromosome. The resulting estimates are shown in **Supplementary Table 5**. They indicate that the X-chromosome results are not an outlier when compared to autosomes with a comparable effective number of loci. For example, the average effective number of independent loci on the X chromosome (1,309) is similar to our estimates for chromosomes 19-22, where the mean estimate is 1,401 (range 1,148 to 1,453). On these autosomes, the number of

lead SNPs is 8.5 (range 8 to 12) compared to our finding of 12 lead SNPs on the X chromosome.

A third factor that is consistent with fewer identified genome-wide significant lead SNPs on the X chromosome compared to autosomes is the combination of haploidy in males and (partial) X-inactivation in females. Both of these lead to a smaller amount of variance explained for a given per-allele effect size when the dosage compensation is not too high, and power of detection is proportional to the proportion of variance explained by a SNP. Note, however, that finding fewer genome-wide significant lead SNPs is also consistent with a smaller per-allele effect size for SNPs on the X chromosome relative to the autosomes, and we are unable to distinguish between these potential explanations in our data.

In a final analysis, we compared the heritability due to common SNPs on the X chromosome to the per-chromosome SNP heritabilities of the autosomes. In this analysis, we used equation (4.1) to estimate SNP heritabilities. We found that relative to autosomes similar in length, the X chromosome has a lower SNP heritability, but once again, the X chromosome has a similar SNP heritability to that of autosomes similar in effective number of loci.

Despite the similarity of the X chromosome to the autosomes with a similar effective number of loci, a comparison to autosomes of a similar length may be more appropriate. This is because, under the common assumption that all SNPs explain an equal amount of heritability in expectation (Bulik-Sullivan et al., 2015; Yang et al., 2011; Vilhjálmsson et al., 2015; Turley et al., forthcoming), the SNP heritability of a chromosome should scale linearly with the number of SNPs on the chromosome. This can be seen in the short proof below.

Standardize the phenotype to have mean zero and variance one. Let  $h_k^2$  denote the SNP heritability for chromosome  $k$ ,  $x_{ij}$  denote the genotype of individual  $i$  at SNP  $j$ , and  $\beta_j$  denote the effect of SNP  $j$  on some phenotype conditional on the genotypes of all other SNPs. We let  $K$  denote the set of SNPs on chromosome  $k$  and  $M_k$  denote number of SNPs on that chromosome. Because we assume that each SNP contributes equally to the heritability of the phenotype in expectation,  $\text{Var}(x_{ij}\beta_j)$  is constant across SNPs. We let  $\sigma_h^2$  denote this constant value. Under the assumption that effect sizes are uncorrelated across SNPs, we calculate:

$$h_k^2 = \text{Var}\left(\sum_{j \in K} x_{ij}\beta_j\right)$$

$$\begin{aligned}
&= \sum_{j \in K} \text{Var}(x_{ij}\beta_j) \\
&= \sum_{j \in K} \sigma_h^2 \\
&= M_k \sigma_h^2.
\end{aligned}$$

Thus, the SNP heritability for chromosome  $k$  is proportional to the number of SNPs on chromosome  $k$ .

On the other hand, under the assumption that the fraction of heritability explained by a SNP is proportional to its LD Score—which is the assumption underlying LDAK (Speed et al., 2017)—it is true that the SNP heritability would be proportional to the effective number of SNPs on the chromosome rather than the total number of SNPs. Given the apparent linear relationship between the SNPs'  $\chi^2$ -statistics and their LD Scores observed in Supplementary Figure 2—consistent with the model underlying LD Score regression—we think it is unlikely that the LDAK assumption holds for *EduYears*. An explanation of the reduced heritability that is more consistent with the observed data is partial dosage compensation through X-inactivation.

**Supplementary Figure 6** summarizes the results from our comparative analyses with results from autosomal analyses.

## 5. Biological Mechanisms

### **5.1. Introduction**

In this section we describe our insights into the biological mechanisms through which genetic variation affects *EduYears*. **Supplementary Figure 7** provides a flowchart detailing our analyses intended to identify enriched tissues/cell types, enriched gene sets, causal genes, the points at which the causal genes are expressed across development, and causal SNPs.

To preview, in broad outline the results have two main themes:

1. The setup of the brain during prenatal development and in the first years after birth; and
2. Online neuronal communication and synaptic plasticity occurring throughout life.

The first theme was very prominent in an earlier GWAS of *EduYears* in a much smaller sample<sup>1</sup> and is even stronger here. The second theme now comes to the foreground as well (**Supplementary Figure 8**).

### **5.2. Methods: Enriched Tissues/Cell Types, Enriched Gene Sets, Causal Genes, BrainSpan Developmental Transcriptome**

*DEPICT*. We used *DEPICT*<sup>45</sup> (downloaded February 2016 from <https://github.com/perslab/depict>) to identify the tissues/cell types where the causal genes are strongly expressed, detect enrichment of gene sets, and prioritize likely causal genes. (A *gene set* is a group of genes annotated as sharing an important property, such as the participation of their products in a common pathway.) We ran *DEPICT* as described previously<sup>1</sup> with the following exceptions:

1. We used 37,427 human Affymetrix HGU133a2.0 platform microarrays to assess whether genes in associated loci are highly expressed in any tissue/cell type<sup>45</sup>.
2. We discarded gene sets that were not well reconstituted<sup>46</sup>. The criterion of poor reconstitution was a failure of the original members of the binary gene set to show reconstituted membership scores significantly different from those of all other genes (Mann-Whitney test,  $P < 0.01$ ).
3. To accommodate the much greater number of lead SNPs, we relaxed the significance threshold for defining a matching SNP in the simulated null GWAS from  $5 \times 10^{-4}$  to  $5 \times 10^{-3}$ .

After mergers, the 3,844 lead SNPs meeting the developer-recommended threshold of  $P < 10^{-5}$  define 1,742 distinct loci (**Supplementary Table 7**). The mean locus length is ~304 kb. The loci altogether cover more than 529 Mb, about 18 percent of the human autosomal genome.

A gene is missing from the DEPICT inventory if it lacks high-quality Affymetrix expression data in the Gene Expression Omnibus. We added any non-DEPICT protein-coding gene with a status of **known** in GENCODE (downloaded February 26, 2015 from <https://www.genencodegenes.org/releases/26lift37.html>) to **Supplementary Table 8** if it either encompasses one of the lead SNPs in a DEPICT-defined locus or has the start site closest to such a SNP. There are some SNPs in the *EduYears* meta-analysis, most of them featuring a rare allele, that are not present in the DEPICT files; we did not attempt to provide biological annotation of any such SNPs.

We found a number of tissues/cell types and reconstituted gene sets in the DEPICT output that are exact duplicates despite having different identifiers. We inspected every instance of duplication in our tissue results. Because there are 10,968 gene sets in the current version of DEPICT, we only inspected duplications of gene sets attaining statistical significance (defined by  $FDR < 0.05$ ) or those that are *least* implicated by our GWAS results (defined by having the highest possible  $P$  value,  $P = 1$ ). Without exception we found that multiple objects with the same content but different identifiers have names with highly similar biological meanings (e.g., **voltage-gated channel activity** and **voltage-gated ion channel activity**).

For each group of tissues/cell types with identical vectors of expression scores, we dropped all but one from **Figure 3A** and **Supplementary Table 6** in the following way. If one of the MeSH terms in the group is a substring of the others, we chose its tissue/cell type for inclusion. In the absence of such a simple relationship within the MeSH tree, we attempted to retain the name with the most general reference (e.g., **lymphatic system** over **lymphoid tissue**). Of the 209 tissues/cell types in the DEPICT inventory, 29 were excluded for being a duplicate.

For each group of significant gene sets with identical vectors of membership scores, we dropped all but one from **Supplementary Table 8** in an analogous manner. We attempted to retain the name with the most general or inclusive reference (e.g., **channel activity** over **passive transmembrane transporter activity**), although we often could not make this distinction and sometimes actively override the choice resulting from it. For example, **neuron spine** is arguably more general than **dendritic spine**, but we retained the latter because to our knowledge spines are found exclusively on dendrites. When filtering gene sets from the InWeb database, we preferred names referring to

gene/protein symbols over Ensembl identifiers and current Ensembl identifiers over retired ones. Of the 1,968 reconstituted gene sets attaining statistical significance in the unprocessed DEPICT output, 61 were excluded for being a duplicate<sup>o</sup>.

The relatively small number of duplicate gene sets cannot substantially affect the FDR calculated by DEPICT. The tissues/cell types do raise some concern. We calculated the FDR associated with each of the remaining unique tissues/cell types using the DEPICT-calculated *P* values as input to the Benjamini-Hochberg procedure; the output produced no conflicts with the DEPICT-calculated FDR.

Many gene sets returned by DEPICT as significant are highly correlated (e.g., the GO, KEGG, and Reactome instances of **axon guidance**) and thus do not represent independent biology. To facilitate the interpretation of the results, as is standard<sup>45</sup>, we applied the Affinity Propagation algorithm<sup>47</sup> to segregate the gene sets into clusters and name each cluster after an exemplary member. The input to the algorithm consisted of the correlations between gene sets over just those genes prioritized by DEPICT in the sense of achieving FDR < 0.05. We used the default settings of the *apcluster* function in R.

We remain interested in how the output of DEPICT's gene-set enrichment analysis varies as a function of the trait studied in the GWAS. We therefore updated Supplementary Table 4.5.3 of an earlier GWAS of *EduYears* in a smaller sample<sup>1</sup>, to reflect the most recent *EduYears* results and new results from studies of other traits, namely migraine<sup>48</sup>, height<sup>49</sup>, and coronary artery disease<sup>50</sup>.

In our previous work, we took the expression data from the BrainSpan Developmental Transcriptome<sup>51</sup> and calculated the average expression in the brain of DEPICT-prioritized *EduYears* genes as a function of developmental stage (using the stage definitions in the original paper). We found that the expression level declines as development proceeds from the early fetal period to adulthood, supporting a predominant role of these genes in prenatal brain development<sup>1</sup>. The number of genes significantly prioritized by DEPICT in the current study is 1,838—a greater-than-tenfold increase. When the mean expression in the brain of our larger collection of genes is plotted as a function of developmental stage, the trajectory is now flat; if the outlying **early childhood** stage is ignored, the mean expression in  $\log_2(1 + \text{RPKM})$  is  $\sim 3$  regardless of the developmental stage (**Figure 3B**). This suggests that newly significant SNPs in the current meta-analysis of *EduYears* often act through causal mechanisms involving postnatally expressed genes.

---

<sup>o</sup> Reconstituted gene sets that are duplicates in DEPICT are not necessarily duplicates in the most current version of their original database. Conversely, reconstituted gene sets that are distinct in DEPICT may currently be identical in their original database, although this case seems less likely.



Given that some of the relevant biological mechanisms likely act at different times, we calculated a specific trajectory for each significantly enriched gene set. We followed all steps described in Supplementary Information section 4.7 of Okbay et al.<sup>1</sup> with the following exceptions:

1. To facilitate comparisons across developmental stages, we only included the six brain regions donated at all 12 developmental stages: the amygdala, hippocampus, inferolateral temporal cortex, anterior cingulate (medial prefrontal) cortex, orbitofrontal cortex, and ventrolateral cortex.

2. Previously, we computed the median expression level in  $\log_2(1 + \text{RPKM})$  of all DEPICT-prioritized genes for each combination of donor and brain region and then each individual's mean of the median expression levels. This time, for each stage, we computed the mean expression level of a given gene over all donors who contributed a particular brain region and then the mean of the regional means. The first procedure leads to a measure of how abundantly a collection of genes was transcribed in a single individual, whereas the second procedure yields a measure of how much a particular gene was transcribed on average in a group of individuals. The first procedure simplifies significance testing; each stage's aggregated data represents a sample of individuals, and we can study what might have been observed in a different such sample. In our current applications, we are less concerned with significance testing and sometimes need to judge whether an individual gene qualifies as “prenatal” or “postnatal.” For these reasons we adopted the second procedure in this work<sup>P</sup>.

3. We weighted the contribution of the  $i$ th DEPICT-prioritized gene to a given stage mean in trajectory  $j$  by  $-\log \int_{Z_{ij}}^{\infty} \varphi(x) dx$ , where  $\varphi$  is the density of a standard normal distribution and  $Z_{ij}$  is the membership score of gene  $i$  in gene set  $j$ .<sup>45,46</sup> Those prioritized genes that most strongly drive the significant enrichment of the gene set were thus given the largest weights in the determination of the trajectory.

The results of calculating the set-specific trajectories in this way are shown in **Supplementary Figure 22** and **Supplementary Table 8**. The weighting scheme, albeit *ad hoc*, does capture a prominent dimension of variation across the enriched gene sets. When we summarized each weighted trajectory as the mean of the prenatal stages minus the mean of the postnatal stages, we found a correlation of 0.88 between the prenatal-postnatal differences of the exemplary gene sets and their projections on the first principal component computed from their correlation matrix.

---

<sup>P</sup> The one exception to this statement is the calculation of the trajectories and their confidence intervals in **Figure 3A**. Here we reverted to our previous procedure.

*Stratified LD Score* regression. We used stratified LD Score regression to supplement our analyses of tissues/cell types. Finucane et al. have made available stratified LD Scores based on whether a gene is highly expressed in astrocytes, oligodendrocytes, or neurons<sup>52</sup>. These annotations are in turn based on the expression data gathered from postnatal mice brains by Cahoy et al.<sup>53</sup> We will provide more details about our use of stratified LD Score regression in the next subsection.

### **5.3. Methods: Robustness Checks of Causal Genes and Enriched Gene Sets**

*MAGMA*. We also employed the tool Multi-Marker Analysis of Genomic Annotation (MAGMA)<sup>54</sup> (downloaded June 28, 2017 from <https://ctg.cncr.nl/software/magma>) for the purpose of gene prioritization. We used the “multi=snp-wise” option, which aggregates a gene-level test of mean SNP association equivalent to VEGAS<sup>55</sup> and a test of the single maximally associated SNP. We mapped a SNP to a gene if it resides within the gene boundaries or 5 kb of either endpoint, according to the coordinate file available at the MAGMA webpage. We used the Europeans in 1000 Genomes phase 3 as the reference panel for estimating LD. We applied the Benjamini-Hochberg procedure and declared a gene to be significant if its joint  $P$  value falls below the threshold corresponding to  $FDR < 0.05$ .

An importance difference between the gene-prioritization functions of DEPICT and MAGMA lies in their treatment of the relationship between SNP and gene. In brief, DEPICT takes a local maximum of the Manhattan plot clearing the threshold  $P < 10^{-5}$  recommended by the DEPICT developers and constructs an LD-based locus centered on this lead SNP. Any gene overlapping this locus is prioritized if its vector of memberships in the DEPICT reconstituted gene sets is significantly correlated with the vectors belonging to genes near other lead SNPs. A prioritized gene is not necessarily the closest to a lead SNP, and in fact the distance between lead SNP and prioritized gene can reach hundreds of kilobases. MAGMA, on the other hand, tests the significance of the SNPs falling directly within the gene boundaries or sufficiently close to one of its endpoints. (The developers recommend using a short radius.) This difference between DEPICT and MAGMA in the treatment of SNP-gene distance is perhaps not particularly consequential because a SNP residing inside a gene can show strong association with the trait as a result of LD with causal SNPs outside the gene. Nevertheless, it is an important conceptual distinction.

*PANTHER*. As a robustness check of the DEPICT results, which are based on its reconstituted gene sets, we use an enrichment analysis that employs binary gene sets (i.e., gene sets for which any given gene is either a member or not, as opposed to having a quantitative measure of the amount of membership). Specifically, we used the

PANTHER binomial overrepresentation test<sup>56,57</sup>, which has been implemented as a web-based tool (<http://www.geneontology.org>). The input to this method is a discrete list of genes supplied by the user; we used all DEPICT-prioritized genes as input. There is no circularity in this procedure, despite the strong overlap in gene sets between DEPICT and PANTHER, because DEPICT's gene-prioritization algorithm relies on correlations between rows of the *gene*  $\times$  *gene set* matrix and not the labels of the matrix columns that drive whatever significant correlations there may be. To verify this, we reran the analysis with all genes in DEPICT-defined loci (including genes absent from the DEPICT inventory) rather than just the prioritized genes and also all MAGMA-prioritized genes. These three input lists produced similar results, although naturally the significant results from the longer lists are fewer and weaker in effect size (results not shown).

The null hypothesis in the PANTHER binomial test is that the input gene list is a random sample of all genes in the reference gene list. There are a number of reasons, however, why rejection of this null hypothesis might not be indicative of true enrichment. For instance, longer genes are inherently more likely to obtrude into GWAS loci, and genes strongly expressed in the brain also tend to be longer. Hence a list of all genes overlapping our DEPICT-defined loci might contain an above-chance number of genes with neural functions even if our trait is not mediated by the brain at all. For this reason we used the genes prioritized by DEPICT in the 2014 GWAS of height<sup>58</sup> as a negative control. (DEPICT was not used in the more recent GWAS of height<sup>49</sup> to prioritize likely causal genes.)

Note that the versions of the binary gene sets used in our application of PANTHER are more recent than the ones employed in the DEPICT reconstitution procedure. In fact, PANTHER incorporates updates of Gene Ontology (GO) on a roughly monthly basis. Also, many of the PANTHER gene sets are not in the DEPICT inventory because of the decision to reconstitute only those gene sets with at least 10 and no more than 500 members at the time<sup>46</sup>. Furthermore, some gene sets have been retired and new ones created. We used the default Bonferroni correction to adjust the PANTHER *P* values for each annotation dataset.

*Stratified LD Score regression.* A principled form of enrichment analysis is to partition the heritability of the trait between SNPs in or near genes that are members of a given set and all other SNPs. This approach is not statistically powerful, however, for two main reasons. The first is that many SNPs in or near a causal gene probably have negligible impact on the trait, which implies that enrichment effect sizes in this type of analysis will tend to be small. The second is that most binary gene sets have few members and hence a relatively small number of SNPs mapping to them; the consequence is large standard errors in a heritability partition. We took these

considerations into account when planning our enrichment analysis with stratified LD Score regression<sup>24</sup>. The developers of LD Score regression have gathered together a number of SNP-level annotations (<https://data.broadinstitute.org/alkesgroup/LDSCORE>), but the method is not restricted to these particular annotations. We devised three types of novel annotations in our gene-set enrichment analysis.

1. We classified a SNP as a member of a gene set if it is located within the boundaries of a gene ranking in the top 10 percent of the DEPICT reconstituted version of that set or within 100 kb of such a gene<sup>q</sup>. The values of 10 percent and 100 kb are taken from a recent work that used the DEPICT tissue/cell type data in an analogous way and found that these settings led to the smallest enrichment *P* values<sup>52</sup>. (The largest enrichments reported in this paper are about 1.4, bearing out our earlier point about small effect sizes.) Since stratified LD Scores are tedious to compute, we examined only the exemplary gene sets chosen by the Affinity Propagation algorithm in the DEPICT analysis pipeline.

2. We also constructed set-specific annotations indicating whether a SNP falls within or less than 100 kb from a member of the original binary gene set in the DEPICT inventory. We chose original gene sets with 200 or more members at the time of the DEPICT reconstitution and whose reconstituted versions are significantly enriched according to the DEPICT analysis. The cutoff of 200 genes comes from a recommendation in live tutorials given by the LD Score regression developers—made with the issue of statistical power in mind—to restrict this type of analysis to annotations borne by at least 1 percent of all SNPs where both alleles are common in the 1000 Genomes European populations.

3. An attraction of heritability partitioning as a means of gene-set enrichment analysis is that it provides a means of assessing whether it is prenatal or postnatal processes that tend toward larger effect sizes. For instance, if **forebrain development** shows greater enrichment than **regulation of synaptic transmission**, we might tentatively conclude that early brain development has a larger genetically mediated impact on *EduYears* than online neurophysiological function.

We took this notion to one possible logical endpoint by classifying all protein-coding genes with data in the BrainSpan Developmental Transcriptome as **non-brain**,

---

<sup>q</sup> The 10-percent cutoff is not intended to imply that a gene set defined by a protein complex such as **npBAF complex** has more than a thousand cryptic subunits encoded by different genes. When a gene becomes a high-ranking member of a gene set upon application of the DEPICT reconstitution procedure, this simply means that the gene follows the same pattern of co-expression as the original members of the set.

**brain/flat**, **brain/prenatal**, and **brain/postnatal** in the following way. At any given temporal stage, the median protein-coding gene has an expression in the brain of roughly 1.8 (on the scale used in **Figure 3B** and **Supplementary Figure 22**). We annotated a gene with an average expression in the brain over all 12 stages smaller than 1.8 as **non-brain**. The remaining genes were then ranked by prenatal-postnatal difference; the genes in the top third were annotated as **brain/prenatal**, those in the middle third as **brain/flat**, and those in the bottom third as **brain/postnatal**. All SNPs inside or within a 100-kb radius of a gene bearing a given annotation inherited the annotation for purposes of this heritability-partition analysis.

When the **brain/prenatal** genes were given as input to the PANTHER binomial test, the highest-ranking GO biological processes by fold enrichment with the strings “neur” and “brain” in their identifiers were **neural tube closure** and **forebrain development**, respectively. The string “synap” did not appear in any of the nominally significant results. When the **brain/postnatal** genes were given as input, **regulation of neuronal synaptic plasticity**, **regulation of synaptic vesicle transport**, and **regulation of neurotransmitter secretion** were three of the top four results. We interpret this pattern as validating our use of the BrainSpan data to classify genes in this way.

In the present work, we calculated the stratified LD Score of each HapMap3 SNP (over all 1000 Genomes SNPs with European MAF > 0.0013 within the recommended 1-cM window) with respect to the gene-set annotations. We then added each gene-set annotation in turn to the baseline set of annotations and regressed the GWAS  $\chi^2$  statistics of the HapMap3 SNPs on the stratified LD Scores. A key innovation since the application of stratified LD Score regression in Okbay et al.<sup>1</sup> is the addition of various baseline annotations referring to properties such as allele age, MAF, and LD<sup>59</sup>. A potential confounder in LD Score regression is heritability per SNP, which clearly affects the  $\chi^2$  statistic and varies as a function of the LD Score itself (**Supplementary Table 31**). By taking into account LD and various related properties, the baseline annotations should now be even more robust. As input to the determination of the regression weights, we used the non-stratified LD Scores of HapMap3 SNPs outside the HLA region (<https://github.com/bulik/ldsc/wiki/Partitioned-Heritability>).

*Non-Significant Gene Sets and Genes.* If all of the null hypotheses in the DEPICT testing of gene-set enrichment are correct, then we should expect roughly 100 of them to show  $P = 1$ , the highest possible value. In our real data, however, 1,769 reconstituted gene sets in fact reach  $P = 1$ . Thus, DEPICT-defined lead SNPs fall far away from the high-ranking members of these gene sets more often than expected by chance. Because it is of interest to know which biological processes are impoverished as well as enriched with respect to *EduYears* lead SNPs, we clustered the  $P = 1$  gene sets with the Affinity

Propagation algorithm in the same way that we clustered the  $\text{FDR} < 0.05$  gene sets of primary concern. We employed the exemplary gene sets chosen in the clustering of the  $P = 1$  results to guide our interpretation of the biology that is minimally involved in the genetic etiology of *EduYears*.

We also employed stratified LD Score regression to estimate the *EduYears* heritability accounted for by SNPs in or near genes that are high-ranking members of the exemplary  $P = 1$  gene sets. We used the same parameter values (top 10 percent, 100-kb radius) as in the analogous application to the  $\text{FDR} < 0.05$  gene sets. Genes expressed strongly in non-CNS tissues show an average *EduYears* heritability enrichment factor of almost exactly unity<sup>52</sup>; although this result can be taken as evidence of stratified LD Score regression being well calibrated, it also suggests that analysis of impoverished gene sets is likely to have poor statistical power. Nevertheless, we hypothesized that these gene sets exhibiting very high  $P$  values in the primary DEPICT analysis will show an average enrichment factor somewhat smaller than unity. We also applied this heritability-partition technique to the original binary gene sets with at least 200 members whose reconstituted versions reach  $P = 1$ .

#### 5.4. Methods: Causal SNPs

*Stratified LD Score regression.* We used stratified LD Score regression<sup>24</sup> to estimate which types of SNPs are most likely to have relatively large effects. We expanded upon previous analyses by incorporating the 450 annotations gathered together by Pickrell<sup>60</sup>. We will call these the “fgwas annotations,” and they are available at <https://github.com/joepickrell/1000-genomes><sup>r</sup>. It is important to keep in mind that the effect sizes of annotations whose instances describe relatively short, highly functional regions of the genome (e.g., **conserved** or **nonsynonymous**) can be much greater than those in the applications of stratified LD Score regression to tissues/cell types and gene sets described earlier because the latter types of annotations typically apply to much longer regions of the genome that are likely to contain many SNPs with negligible impact on the phenotype.

---

<sup>r</sup> Okbay et al.<sup>1</sup> used the annotations accompanying the LD Score regression software that contain references to distinct tissues. Each of these annotations, however, is formed by taking a union of SNPs associated with assayed histone marks across a variety of cell types and developmental stages. The SNPs that bear one of these annotations may be in fact quite heterogeneous, and thus the fgwas annotations offer the potential of greater resolution. Many of the same data sources contribute to both the LD Score regression and fgwas annotations, and the latter can therefore be viewed to some extent as a disaggregation of the former. We conduct stratified LD Score regression rather than using the fgwas tool<sup>60</sup> itself because simulations suggest that stratified LD Score regression will usually have greater power to detect enrichment<sup>24</sup>.

We calculated the stratified LD Scores and performed the regressions in the same manner as described in previous subsections.

*CAVIARBF*. For fine mapping, we use the tool *CAVIARBF*<sup>61,62</sup>. This program has a couple of advantages over other fine-mapping methods. First, it offers a systematic method of jointly incorporating multiple annotations. It does this by imposing a regularization penalty on the estimated effects of the annotations. This allows *CAVIARBF* to use LASSO ( $L_1$ ), ridge ( $L_2$ ), or elastic net (combination of  $L_1$  and  $L_2$ ) penalties<sup>63</sup> to select annotation sets for inclusion in the model. The optimal penalty parameter can be selected using various model-selection methods (e.g., cross validation). Second, *CAVIARBF* allows for more than one causal SNP per locus<sup>s</sup>.

We used annotations from several sources. We used the 74 baseline annotations employed by stratified LD Score regression<sup>64</sup>, which include the 52 functional annotations from the original paper<sup>24</sup>. We combined these 74 annotations with annotations from fgwas<sup>60</sup>. We thus have 525 annotations in total, of which 8 are quantitative annotations. We applied a MAF filter of 0.01 and a sample-size filter of 400,000. This resulted in a set of 7,951,231 SNPs; two genome-wide significant SNPs were filtered out (due to having MAF below 0.01).

To make it computationally tractable to allow for multiple causal SNPs within a locus, we only considered SNPs near lead SNPs clearing the genome-wide significance threshold  $P < 5 \times 10^{-8}$ . As in the *CAVIARBF* paper<sup>62</sup>, we only used SNPs within a 100-kb window (50-kb radius) of these lead SNPs<sup>t</sup>. After limiting ourselves to SNPs within 50 kb of a lead SNP, we retained 332,837 SNPs, of which 331,268 have been annotated with our set of 525 labels. The number of loci decreased from 1,269 to 1,067 because lead SNPs with overlapping 100-kb windows were merged.

To estimate pairwise LD between the SNPs in each locus, we use the subset of Europeans in the 1000 Genomes Phase 3 reference panel. To address potential concerns that the small sample size of the 1000 Genomes reference panel will result in inaccurate

---

<sup>s</sup> The results of our gene-prioritization analyses indicate that there are many thousands of genes affecting *EduYears*. Thus, the assumption underlying the tool fgwas<sup>60</sup> that each of  $\sim 1,700$  independent segments of the genome contains at most one causal SNP seems unreasonable. When there is only one causal site in a locus, fgwas and *CAVIARBF* perform similarly in simulations<sup>62</sup>. However, when true number of causal sites per locus is increased, running *CAVIARBF* with up to three causal sites per locus demonstrates better performance.

<sup>t</sup> One concern with this approach is that it may leave out causal SNPs that are not sufficiently close to any GWAS hits. Based on simulations with whole-genome sequencing data, however, a recent paper by Wu et al.<sup>357</sup> found that 89% of the time, the causal SNP is located within 50 kb of the top GWAS hit. A caveat is these authors simulated one causal SNP at a time, whereas actual GWAS results will involve many causal SNPs and GWAS hits at once.

LD estimation and thus errors in fine mapping<sup>65</sup>, we repeat our analyses using a reference panel comprised of ~400,000 Europeans from the UK Biobank (the *UKB* reference sample, as in **Supplementary Section 1.9**).

To compute Bayes factors for each SNP, we needed to define a prior for the variance of effect sizes. As in prior GWAS of *EduYears*<sup>1</sup>, we averaged over prior variances of 0.01, 0.1, and 0.5. We set the following additional parameters for CAVIARBF. The sample size was set to the mean sample size of our *EduYears* meta-analysis: 1,076,358. We computed exact Bayes factors. As suggested in<sup>62</sup>, we added a 0.2 to the main diagonal of the LD matrix because we used a reference panel for LD estimation<sup>u</sup>.

We computed results for the case of at most two causal SNPs per block. (Because CAVIARBF computes Bayes factors for each potential set of causal SNPs within a block, and because each block can contain thousands of SNPs, assuming more than two causal SNPs quickly becomes computationally intractable.) Another important assumption is that the causal sites have indeed been included in each block. This may turn out not be the case if the causal site has been excluded by a QC filter or is a type of polymorphism (e.g., an indel) that has not studied in our meta-analysis.

CAVIARBF provides a number of options for handling annotations, and we used elastic net regularization. This should be the best option when there are many annotations of some value (**Supplementary Tables 37 and 10**) and when overfitting needs to be avoided. The elastic net parameters  $\alpha$  and  $\lambda$  (as defined in the relevant CAVIARBF paper<sup>62</sup>) were selected via 5-fold cross-validation, from all combinations of  $\alpha \in \{0, 0.5, 1\}$  and  $\lambda \in \{2^{-2}, 2^{-1}, \dots, 2^{10}\}$ . This procedure includes both LASSO and ridge regression as special cases;  $\alpha = 0$  gives ridge regression (an  $L_2$  penalty equivalent to a normal prior on annotation effect sizes);  $\alpha = 1$  gives LASSO (an  $L_1$  penalty equivalent to a Laplace prior); and  $\alpha = 0.5$  places equal weight on the  $L_1$  and  $L_2$  penalties.

Results are reported as posterior inclusion probabilities (PIPs), which are defined for each SNP considered. The PIP for a SNP gives the posterior probability that a SNP is causal given the GWAS data, annotation data, and estimated annotation coefficients<sup>v</sup>.

---

<sup>u</sup> An additional concern emerges when (1) the LD matrix is estimated in a separate reference sample or (2) the sample size differs nontrivially across SNPs (perhaps as a result of QC filters). Under these conditions, some SNPs with slightly different Z-statistics in our meta-analysis may be in perfect LD<sup>62</sup>. When this occurs, CAVIARBF will tend to assign a much higher posterior inclusion probability to the SNP with the slightly higher Z-statistic, although intuitively there should not be any difference between the SNPs.

<sup>v</sup> We obtained a nearly identical set of SNP candidates clearing  $\text{PIP} > 0.9$  using the topK method. For example, the SNP rs61734410—which we highlight later because of its nonsynonymous status and residence within a gene encoding a voltage-gated calcium channel pore-forming subunit—is prioritized by both topK and elastic net. Interestingly, it is CAVIARBF’s ability to allow for multiple causal SNPs in a locus rather than the choice of annotation method that is critical to the prioritization of this SNP. The



### 5.5. Results: Enriched Tissues/Cell Types

Without exception, all 23 tissue/cell types with the MeSH first-level term **nervous system** were found to be statistically significant by DEPICT in the sense of  $FDR < 0.01$  (**Figure 3A** and **Supplementary Table 6**). Under certain conditions, some tissues/cell types in this dataset have been shown to be susceptible to false positives when analyzed with DEPICT<sup>52</sup>, but none of these falls under the MeSH second-level term **central nervous system**.

We will discuss the top tissues/cell types in order of  $P$ -value ranking (although we caution against taking the ranking itself too seriously). The most significant result is **hippocampus** ( $P = 1.87 \times 10^{-24}$ ), which in humans is believed to be responsible for the formation of long-term memories. The second and fourth most significant results, **brain** and **central nervous system** respectively, are too high in the MeSH tree to provide much insight. The third most significant result is **limbic system** ( $P = 2.78 \times 10^{-24}$ ), which refers to a collection of functionally heterogeneous regions that includes the hippocampus.

The fifth most significant result is **cerebral cortex** ( $P = 3.92 \times 10^{-24}$ ), which refers to the layer of gray matter making up the surface of the brain. In humans, it is dominated by the evolutionarily recent *neocortex*, which is the seat of higher mental functions such as perception, thought, language, spatial visualization, and the initiation of voluntary movement.

Some significant tissues/cell types fall outside the first-level MeSH term **nervous system**. The first of these is **retina** ( $P = 2.05 \times 10^{-16}$ ), which is perhaps not anomalous; the retina is essentially an outgrowth of the embryonic forebrain. The retina contains layers of neurons, whose connections to the rest of the central nervous system through the optic nerve are laid down as a result of synaptogenetic mechanisms (involving semaphorins, ephrins, and the ROBO receptor) highlighted in previous biological annotation of *EduYears* GWAS results<sup>1</sup>. The cell type **neural stem cells** ( $P = 1.71 \times 10^{-10}$ ) falls under the first-level term **cells** but is nonetheless obviously neural.

We note that, despite the large number of prioritized genes, many tissue/cell types are *not* prioritized. In fact, 93 of the 157 non-neural tissues/cell types show  $P > 0.95$  (**Figure 3A** and **Supplementary Table 6**).

At the level of distinct cell types, not all neural cell types are equal. Our application of stratified LD Score regression showed that whereas SNPs in or near genes highly

---

LocusZoom plot (**Supplementary Figure 9C**) and other analyses (results not shown) indicate that there is at least one additional causal SNP in this gene driving most of the association signals.

expressed in neurons account for a significantly enriched share of heritability (1.33-fold,  $P = 2.89 \times 10^{-11}$ ), genes highly expressed in astrocytes (1.08-fold,  $P = 0.07$ ) and oligodendrocytes (1.09-fold,  $P = 0.06$ ) do not show significant enrichment even with our massive sample size (**Supplementary Table 9**). Moreover, the partial regression coefficients of these two glial-cell annotations are both very close to zero ( $\tau = -2.78 \times 10^{-10}$ ,  $P = 0.79$ ;  $\tau = 1.64 \times 10^{-11}$ ,  $P = 0.88$ ).

To further examine results related to glial cells, we examined both the significant and non-significant results of our primary DEPICT gene-set enrichment analysis. The significantly enriched gene sets are listed in **Supplementary Table 8**. One cluster of gene sets is named after **regulation of gliogenesis**. Four of this cluster's members are **negative regulation of glial cell differentiation** ( $P = 7.31 \times 10^{-3}$ ), **regulation of glial cell differentiation** ( $P = 2.09 \times 10^{-3}$ ), **negative regulation of gliogenesis** ( $P = 3.04 \times 10^{-3}$ ), and **regulation of gliogenesis** ( $P = 1.82 \times 10^{-3}$ ) itself.

In the binary version of GO, **negative regulation of glial cell differentiation** is a subset of **regulation of glial cell differentiation**. The definition of **negative regulation of glial cell differentiation** is “[a]ny process that stops, prevents, or reduces the frequency, rate or extent of glia cell differentiation” (<http://amigo.geneontology.org/amigo/term/GO:0045686>, accessed August 4, 2017). Similarly, the binary **negative regulation of gliogenesis** is a subset of the binary **regulation of gliogenesis**, and the definition of the former is “[a]ny process that stops, prevents, or reduces the frequency, rate or extent of gliogenesis, the formation of mature glia” (<http://amigo.geneontology.org/amigo/term/GO:0014014>, accessed August 22, 2017).

Because neural progenitors tend to become neurons when they exit the cell cycle earlier and glial cells when they do so later, the significance of **negative regulation of glial cell differentiation** and **negative regulation of gliogenesis** plausibly supports the greater relative importance of neurons over glial cells. Also, whereas both **positive regulation of neurogenesis** ( $P = 1.56 \times 10^{-5}$ ) and **negative regulation of neurogenesis** ( $P = 2.60 \times 10^{-5}$ ) are significant in our results, **positive regulation of glial cell differentiation** is not ( $P = 0.58$ ). (**Positive regulation of gliogenesis** is not present in the current DEPICT inventory of gene sets.)

The pattern is similar in the case of gene sets specifically defined by one of the two glial-cell types. None of the three gene sets with the string “astroc” in its identifier is statistically significant at the  $FDR < 0.05$  level (although they all satisfy  $P < 0.11$ ). None of the five gene sets with the string “oligodend” in its identifier is statistically significant either; the closest are **regulation of oligodendrocyte differentiation** and **negative**

**regulation of oligodendrocyte differentiation** (both  $P = 0.02$ ). Of the eight gene sets with the string “myelin” in its identifier, the lowest  $P$  value is 0.24, and **central nervous system myelination** has  $P = 0.85$ .

It is perhaps surprising that SNPs mapping to genes highly expressed in oligodendrocytes fail to account for a significantly enriched share of *EduYears* heritability—especially in light of the fact that reaction time in simple cognitive tasks is negatively correlated with cognitive performance<sup>66–69</sup>. After all, myelination renders the action potential saltatory and thereby increases its velocity along the axon by an order of magnitude, which might suggest the hypothesis that variation in cognitive phenotypes across individuals might be related to genetic variation in genes related to myelination.

## 5.6. Results: Causal Genes and Enriched Gene Sets

**Supplementary Table 7** lists all genes in the DEPICT-defined loci, regardless of prioritization  $P$  value. A total of 1,838 genes were “significantly prioritized” by DEPICT, as defined by having  $FDR < 0.05$ . Of the 1,742 loci, 1,068 contain at least one prioritized gene (61 percent). Multiple genes were prioritized in 312 loci, indicating that a given locus may contain not only multiple causal SNPs but also more than one causal gene.

MAGMA prioritized 8,171 genes according to the criterion of  $FDR < 0.05$  (**Supplementary Table 29**), nearly 45 percent of all genes present in its inventory. Reassuringly, 1,583 of the DEPICT-prioritized genes were also prioritized by MAGMA (86 percent).

A total of 1,907 unique reconstituted gene sets are significantly enriched by DEPICT. The Affinity Propagation algorithm grouped these sets into 143 clusters. **Supplementary Figure 22** shows a subset of the exemplary gene sets (i.e., the gene sets for which the clusters are named) that have an enrichment factor from LD Score regression exceeding 1.25 (see **Supplementary Table 35**); each row of the heat map gives the set-specific mean expression in the brain of the DEPICT-prioritized genes at each of the BrainSpan Development Transcriptome stages. The rows are ordered by prenatal-minus-postnatal effect size. It is striking how well this ordering reflects the known course of brain development. For instance, **npBAF complex** (the prefix “np” stands for “neural progenitor”), **regulation of nervous system development** (this cluster includes many gene sets defined by progenitor proliferation and neurogenesis), **telencephalon cell migration**, **axon guidance (Reactome)**, and **dendrite morphogenesis** are all positioned in the correct order relative to each other. **Supplementary Table 8** gives the trajectory of each individual gene set in numerical form.

To see which functional systems are *least* implicated by the *EduYears* GWAS results, we examined the 1,769 reconstituted gene sets that reached  $P = 1$ , the highest possible  $P$  value. These gene sets were grouped into 237 clusters (**Supplementary Table 27**). The biology represented by these clusters is quite diverse: **enlarged spleen, abnormal trabecular bone morphology, blood vessel development, formation of fibrin clot (clotting cascade), cholesterol transport, telomere maintenance**, and so forth. Some of this biology has been implicated in GWAS of other traits (**Supplementary Table 28**). There are several distinct clusters defined by the immune system (e.g., **decreased T cell proliferation**), and impoverishment of this biology is a recurring theme of our analyses.

A useful feature of PANTHER is that it returns negatively enriched gene sets ( $< 1$ -fold) as well as positively enriched ones (**Supplementary Table 30**). There are several sets defined by immunity (e.g., **immune response**) that are barren of DEPICT-prioritized genes. The sets that do overlap with the DEPICT-prioritized genes strongly bear out the importance of nearly all stages of neuronal development and function. For example, the list of significant GO biological processes is dominated until about position 100 by sets that are not significant in the negative-control analysis of height genes and are clearly defined by the brain (e.g., **neurogenesis, neuron migration, positive regulation of neuron differentiation, central nervous system projection neuron axonogenesis, regulation of dendrite development, regulation of ion transmembrane transporter activity, voltage-gated channel activity, neurotransmitter secretion, synaptic vesicle cycle, glutamate receptor signaling pathway, regulation of long-term synaptic potentiation**). As we discuss below, some of the gene sets shared with height (e.g., **chromatin organization**) are probably not spurious but rather may owe their significance for both traits to the importance of chromatin remodeling in early cell development.

Earlier we discussed the relatively poor enrichment of glial cells, evident in our analyses of both tissues/cell types and gene sets. Another aspect of neurobiology that seems to be little enriched, if at all, is the transport along the microtubules of important molecules between the soma and the neurite tips. The non-significant gene sets returned by DEPICT include the following: **axon cargo transport** ( $P = 0.12$ ), **abnormal axonal transport** ( $P = 0.23$ ), **microtubule-based transport** ( $P = 0.29$ ), **organelle transport along microtubule** ( $P = 0.37$ ), and **vesicle transport along microtubule** ( $P = 0.41$ ). Four of these five gene sets are taken from GO, and none of these appears in the list of significant results returned by PANTHER (**Supplementary Table 30**). In this latter check, we ensured that the absence from the PANTHER results is not the result of the accession number changing its name.

As one of our robustness checks of the primary DEPICT gene-set enrichment analysis, we now turn to the heritability-partition results obtained with stratified LD Score regression. The results obtained with the reconstituted gene sets are shown in **Supplementary Table 35**, which are quite consistent with the primary results. Those gene sets with low DEPICT  $P$  values ( $\text{FDR} < 0.05$ ) show estimated enrichments greater than unity without a single exception, and the vast majority of those sets whose DEPICT  $P$  value is 1 (the highest possible  $P$  value) show enrichments smaller than unity.

The heritability partitions using the original binary gene sets (**Supplementary Figure 25** and **Supplementary Table 36**) are also consistent with the DEPICT gene-set enrichment results. Although the larger standard errors mean that few of the  $P = 1$  gene sets show an enrichment factor significantly different from unity, the average factor over these sets of 0.91 suggestively points to the reduced share of *EduYears* heritability explained by these sets. Many of the  $P = 1$  gene sets with the smallest enrichments are defined by the immune system, even after taking into account set overlap (**Supplementary Figure 25**). Most of the  $\text{FDR} < 0.05$  sets are nominally significant (96 of the 113 such sets exhibiting positive enrichment reach  $P < 0.05$ ), and their average enrichment factor is 1.33. It is perhaps surprising that the very top results are defined by mRNA processing, but we will discuss a plausible potential biological interpretation below.

The Spearman correlation between the heritability-enrichment factors of the binary gene sets in **Supplementary Table 36** and the DEPICT  $P$  values of their reconstituted counterparts in **Supplementary Table 8** is  $-0.49$ . Note that the two methods being compared here—DEPICT on the one hand, stratified LD Score regression with binary gene sets on the other—take markedly different approaches. Nevertheless, these two methods do arrive at reasonably similar rankings of those gene sets initially found to be significant by DEPICT. (It is also of interest that the estimated heritability-enrichment factors of the reconstituted and binary gene sets, where both estimates are available, have a Spearman correlation of 0.68.) Consistent with the lower statistical power of heritability partitioning when applied to an annotation with relatively few SNPs, the mean  $-\log_{10}(P \text{ value})$  of DEPICT exceeds that of stratified LD Score regression with binary gene sets by 3.66.

In summary, our robustness checks affirm our primary DEPICT gene-set enrichment results.

The top results by enrichment effect size in **Supplementary Table 35** are a mix of prenatal and postnatal gene sets (e.g., **protein binding transcription factor activity** vs. **synapse part**). We can similarly characterize **Supplementary Table 36** (e.g., **RNA**

**splicing** vs. **synapse part**). To address whether there is a clear difference in the contribution of processes taking place before birth rather than after, we turn to the heritability-partition results in **Supplementary Table 33**. Those genes annotated by ourselves (using the BrainSpan Developmental Transcriptome data as described above in Section 5.3 under “Stratified LD Score regression”) as **brain/prenatal** show a heritability enrichment of 1.27 ( $P = 2.14 \times 10^{-10}$ ), whereas those genes annotated as **brain/postnatal** show an enrichment of 1.19 ( $P = 1.04 \times 10^{-8}$ ). If we assume (unrealistically) no sampling covariance between these two estimates, then this difference is not statistically significant. Thus, we cannot say that genes that are more strongly expressed in the brain prenatally explain a clearly greater share of *EduYears* heritability than genes that are more strongly expressed postnatally. A more definitive resolution of this issue will require an even larger sample size and perhaps an improved annotation scheme.

We organize our subsequent discussion of the likely causal genes and significantly enriched gene sets with the aid of **Supplementary Figure 22** and **Supplementary Table 8**, proceeding roughly from the most prenatally active gene sets at the top to the more postnatally active ones at the bottom. We will mention several genes and their products<sup>w</sup>, varying in the level of evidence supporting their causal role in the determination of *EduYears*. We adopt a typeface convention to indicate the level of evidence. We assign each gene a point for each of the following criteria that it meets:

1. prioritization by DEPICT in the sense of achieving a  $P$  value low enough to satisfy  $FDR < 0.05$ ;
2. prioritization by MAGMA, also in the sense of achieving  $FDR < 0.05$ ; and
3. residence in a DEPICT-defined locus with at least one genome-wide significant SNP ( $P < 5 \times 10^{-8}$ ).

The name of a gene or its product will appear in **blue** if it has a score of one, **orange** if it has a score of two, and **red** if it has a score of three. When referring to a channel complex, we will use the color corresponding to its pore-forming subunit.

Our discussion below is necessarily selective. We cannot hope to mention every prioritized gene or significantly enriched gene set, although we tend to follow **Supplementary Figure 22** in that we focus on the sets with the largest effect sizes.

---

<sup>w</sup> Our preference for the most recent symbol adopted by the HUGO Gene Nomenclature Committee at the time of writing sometimes means that the symbol present in our inventories is outdated, and in such cases the reader will have to look up the old symbol in **Supplementary Tables 7** and **29** in order to find more information about the gene. We have attempted to give the old symbol in parentheses whenever there is such a conflict.

*Early Brain Development: Chromatin Modification and Transcription.* Among the first few clusters in **Supplementary Figure 22** are **chromatin modification** and **protein binding transcription factor activity**. Similar clusters of gene sets attained significance in an earlier GWAS of *EduYears*<sup>1</sup>. Our temporal analysis here suggests that they correspond to the earliest developmental events that ultimately affect *EduYears*.

Mechanisms for regulating the compactness of chromatin and the accessibility of regulatory regions to nuclear factors are represented in the group of clusters extending very roughly from the initial rows of **Supplementary Figure 22** to **N-acetyltransferase activity**. The chromatin landscape in neural progenitor cells must suppress the genes responsible for the noncommittal pluripotency of embryonic stem cells while still promoting proliferation; it must also suppress neurogenesis until the appropriate time for exiting the cell cycle. The landscape responsible for this state is shaped by at least two *ATP-dependent chromatin-remodeling complexes*<sup>70,71</sup>. All ATP-dependent chromatin-remodeling complexes contain an ATPase subunit and associated subunits with auxiliary roles in modulating catalytic activity and binding to nucleosomes. The assembly of specific isoforms of the associated subunits allows the resulting version of the complex to be recruited to regions near genes that need to be regulated in a particular cell type. The first chromatin-remodeling complex presented here gives its name to the cluster **npBAF complex**<sup>72,73</sup>. (The prefix “np” stands for “neural progenitor,” to distinguish the complex from its precursor form found in embryonic stem cells and its successor found in neurons.) The distinct forms of the BAF complex arise from combinatorial assembly of component proteins from homologous subunit families, and nine components that can or must occupy their respective slots specifically in the npBAF complex are encoded by our prioritized genes (*ARID2*, *ARID1B*, *SMARCC1*, *SMARCA2*, *BCL11A*, *BCL11B*, *PBRM1*, *PHF10*, *BCL7A*). Once bound to a nucleosome, a chromatin-remodeling complex can expose a regulatory region through a number of mechanisms, including sliding the DNA along the histone octamer or even ejecting the octamer entirely<sup>74</sup>. The npBAF complex in particular can also bind in turn to repressive transcription factors, which physically interact with the exposed regulatory region to inhibit the transcription of the nearby neuronal gene.

There are many examples where perturbing a component of the npBAF complex demonstrably or plausibly alters the proliferation of neural progenitors, including knockdown (overexpression) of *PHF10* reducing (increasing) the number of actively dividing progenitors in mice<sup>72</sup>, knockout of *SMARCC2* (which encodes the successor of *SMARCC1* found in the neuronal BAF complex) increasing the pool of basal progenitors and ultimately enlarging the cerebral cortex in mice<sup>75</sup>, and *de novo* mutations of *ARID1B*<sup>76–78</sup> and *SMARCA2*<sup>79</sup> being frequent causes of intellectual disability

accompanied by *microcephaly* (reduction of head size) or *macrocephaly* (enlargement of the head). The syndromes caused by mutations of *ARID1B* and *SMARCA2*, known respectively as Coffin-Siris and Nicolaides-Baraitser syndrome, are very similar and share many symptoms in common other than microcephaly. *De novo* mutations of *SOX11*, which encodes a transcription factor and is itself regulated by the BAF complex<sup>80</sup>, can also cause Coffin-Siris syndrome<sup>81,82</sup>.

Another group of ATP-dependent chromatin remodelers represented in our prioritized genes is the CHD family (*CHD2*, *CHD3*, *CHD6*, *CHD8*)<sup>83</sup>. *De novo* mutations of *CHD2* have been implicated in intellectual disability accompanied by epilepsy<sup>84,85</sup>, and its knockdown in mice inhibits the regenerative proliferation of neural progenitors and promotes the premature production of neurons<sup>86</sup>. *CHD8* has been particularly well studied because of its role in autism spectrum disorder (ASD)<sup>87</sup>. (ASD shows a positive genetic correlation with both cognitive performance and *EduYears*<sup>14,88</sup> and is often accompanied by macrocephaly<sup>89</sup>, although some of the enlargement may arise postnatally<sup>90</sup>.) Perturbation of *CHD8* affects the expression of many genes, particularly those that are highly expressed in early fetal development, and produces macrocephaly in both zebrafish and mice<sup>91,92</sup>.

Yet another chromatin remodeler represented in our list of genes is the NuRD complex. This is the remodeler that is perhaps structurally and functionally the most poorly understood<sup>93</sup>. One slot in the complex must be filled by either *GATAD2A* or *GATAD2B*; another must be filled by *CDK2AP1*; another can be filled by *MTA2*; yet another, by *CHD3*. *De novo* mutations of *GATAD2B* have been implicated in severe intellectual disability, sometimes accompanied by microcephaly<sup>94,95</sup>.

Another mechanism of chromatin modification is the addition or removal of methyl (CH<sub>3</sub>) or acetyl (COCH<sub>3</sub>) “marks” to the “tails” of the histones H3 and H4. Typically, this modification neutralizes the basic charge of lysine and loosens the chromatin, granting the transcription machinery access to the DNA sequence. In contrast, the process corresponding to **histone deacetylase complex** and **histone deacetylase activity (H3-K9 specific)** removes acetyl groups and thereby prevents transcription. The exemplary gene set **histone methyltransferase activity** is defined by the addition of methyl groups to histones. Although histone methylation more often results in transcriptional repression, it can activate transcription depending on the amino acid that is methylated or the surrounding pattern of histone marks. (Methylation of DNA itself is also an important mechanism of gene regulation. For instance, the NuRD complex can be recruited to methylated CpG sites.) Histone marks can also affect chromatin structure by themselves recruiting histone-marking or chromatin-remodeling complexes, which can in turn establish or remove marks.



*KAT5* and *KAT6B* encode lysine acetyltransferases. *KAT5* does not catalyze the acetylation of histones but rather the ATM kinase, which triggers the repair of the damaged DNA that initially recruited *KAT5* to the chromatin<sup>96</sup>. This ensures the survival of the cell and its progression through mitosis. *DMAP1* encodes a protein that can serve as a subunit or interaction partner of both histone acetyltransferases and deacetylases; in the former role, it is a regulator of ATM<sup>97</sup>. *SSBP2* (formerly *SOSS-B2*) encodes a subunit of a complex binding to single-stranded DNA that also contributes to damage repair and cell-cycle progression<sup>98</sup>. *KAT6B* encodes a subunit of a histone acetyltransferase complex, and *de novo* mutations of the gene are known to cause various syndromic forms of intellectual disability, including some accompanied by microcephaly<sup>99–101</sup>.

The SAGA transcription coactivator complex can affect gene expression in multiple ways through its functionally independent modules<sup>102–104</sup>. *SGF29* (also known as *CCDC101*) encodes a subunit of a module that functions as a histone acetyltransferase, whereas *ATXN7* and *ATXN7L3* encode subunits that together work to remove ubiquitin from histones and other substrates. (We will describe ubiquitin shortly.) *TAF5* and *TAF6* encode core subunits of both SAGA and the transcription factor IID complex; the latter coordinates the activities of numerous other proteins needed for the initiation of transcription by RNA polymerase II<sup>105</sup>. Note that many members of the **protein binding transcription factor activity** cluster (e.g., **RNA polymerase II transcription cofactor activity**) are defined by some type of interaction with the basic transcriptional machinery.

Genes that encode histone methyltransferases are also represented in our results (*SETD2*, *SETDB1*, *KMT5A*). Brain-specific deletion of *SETDB1* (formerly *ESET*) in mice leads to a decrease in proliferation and an increase in cell death, particularly among basal neural progenitor cells destined for the deeper layers of the neocortex<sup>106</sup>. This study observed a slight reduction of the H3K9me3 mark in the developing mouse brain, but the mechanism may involve interaction with PRC2 (a repressor complex) to place H3K27me3<sup>107</sup>. *EZH2* (a subunit of PRC2) and *MTF2* (a transcription factor interacting with PRC2) are also encoded by prioritized genes. *KMT5A* (formerly *SETD8*) is unique among histone methyltransferases in that its levels oscillate during the cell cycle. It has been implicated in DNA repair, the condensation of chromosomes before their replication, and replication itself<sup>108</sup>.

*PHF21A* (also known as *BHC80*) encodes a component of LSD1, a histone demethylase complex that itself binds to unmethylated H3K4 (the “null” mark H3K4me0)<sup>109</sup>. Occupancy of a promoter by LSD1 acts to repress the transcription of neuronal genes, including *SCN3A*<sup>110</sup>, and *de novo* mutations of *PHF21A* have been implicated in a syndromic form of intellectual disability that can be accompanied by

microcephaly<sup>110–112</sup>. Suppression of *PHF21A* in zebrafish causes neuronal cell death and a small-head phenotype<sup>110</sup>.

Besides being part of the cytoskeleton, actin is a component of chromatin remodelers and interaction partner of the RNA polymerases. Manufactured in the cytoplasm by the ribosomes, actin must be rapidly shuttled into the nucleus by the import factor encoded by *IPO9* in order for transcription to proceed at the maximal rate<sup>113</sup>. The nucleus is not the final stop; *XPO6* encodes the export factor that transports actin back to the cytoplasm. These processes define significantly enriched gene sets such as **nuclear import** and **nuclear transport**.

Our prioritized genes include at least eight that encode mitogen-activated protein kinases or interacting proteins (*MAPK7*, *MAPK9*, *MAPKAP1*, *MAPK8IP3*, *MAP2K1*, *MAP2K5*, *MAP3K2*, *MAP3K3*)<sup>x</sup>. Perturbation of *MAPK7* (formerly *ERK5*) has been shown to reduce the number of neurons<sup>114–116</sup>, although the studies appear not to agree about the decision point where the relevant branch is taken, possibly because of variation across model organisms. *MAPK7* can be phosphorylated by *MAP2K5* (formerly *MEK5*), which in turn can be phosphorylated by *MAP3K2* (formerly *MEKK2*) and *MAP3K3* (formerly *MEKK3*)<sup>117,118</sup>.

MAPKs belong to a larger group of kinases called CMGC (named after the initials of some members). One of the families in this group consists of the cyclin-dependent kinases (CDKs), serine/threonine kinases whose activity depends on a regulatory subunit called a cyclin<sup>119,120</sup>. Our prioritized genes include several that encode CDKs (*CDK2*, *CDK4*, *CDK5*, *CDK10*, *CDK12*, *CDK13*, *CDK14*, *CDK19*). Early work on the CDKs established their role in the cell cycle, but they have undergone enormous evolutionary specialization upon divergence and now frequently act as regulators of transcription. *CDK12* happens to forge a link between transcription and the cell cycle; its product phosphorylates RNA polymerase II and is specifically required for the transcription of genes involved in DNA repair<sup>121</sup>.

A *de novo* mutation of *CDK19* has been observed in one patient with mild intellectual disability accompanied by microcephaly<sup>122</sup>. This gene encodes one of the proteins that can serve as the enzymatic component of the Mediator complex, which is required for transcription by RNA polymerase II and serves as a focus of regulatory

---

<sup>x</sup> A *mitogen* is any substance that promotes mitosis, and the first well-characterized members of the MAPK family are indeed activated by mitogens and when inhibited lead to a cessation of proliferation<sup>358</sup>. These proteins are involved in a variety of functions, however, even in a neural context<sup>118,359</sup>. (The exemplary gene set **MAPK targets/nuclear events mediated by MAP kinases** is in fact closer to the bottom of Supplementary Figure 22.)

signals<sup>123</sup>. The Mediator complex also has roles in chromatin modification, mRNA processing, and other aspects of gene regulation. There are several prioritized genes that encode other components of the complex (*MED1*, *MED13*, *MED13L*, *MED19*, *MED21*, *MED26*, *MED27*, *MED28*, *MED30*). *De novo* mutations of *MED13L* cause a syndromic form of intellectual disability that may be accompanied by microcephaly in a minority of patients<sup>124,125</sup>.

*Early Brain Development: Post-Transcriptional Regulation of Gene Expression.* We now move on from the processes controlling transcription and focus on the next checkpoint in gene expression: the regulation of mRNA once the key regions of chromatin have been opened and the gene has been transcribed. Many of the relevant genes are high-scoring members of the reconstituted gene sets giving their names to the clusters **mRNA splicing** and **regulation of nuclear mRNA splicing, via spliceosome**.

Several of our prioritized genes encode recognized splicing factors (*SRSF6*, *SRSF9*, *HNRNPA1*, *HNRNPK*, *KHDRBS3*, *PCBP3*, *PCBP4*, *RBFOX1*, *NOVA1*, *RBM4*, *RBM4B*, *RBM14*)<sup>126</sup>. There are eight additional DEPICT-prioritized genes containing the canonical RNA-binding motif (*RBM5*, *RBM6*, *RBM12*, *RBM15B*, *RBM23*, *RBM27*, *RBM39*, *RBMS3*)<sup>127</sup>. Despite often being among the most strongly expressed genes in the brain<sup>51,128</sup>, many of these genes and others involved in mRNA regulation seem to have been little studied in a neural context (perhaps because they are often ubiquitously expressed). These genes show a range of peak expression times. We focus on those that have received some coverage in the literature and seem to be involved mainly in early prenatal development<sup>129</sup>.

*SRSF6* (formerly called *SRP55*) has been shown to influence the splicing of *HTT* transcript, producing an isoform that is more often found in the brains of Huntington's disease patients<sup>130,131</sup>. A systematic attempt to identify the transcripts targeted by RNA-binding proteins found that *SRSF6* targets enrich GO gene sets that are among our own significantly enriched sets (e.g., **transcription factor binding**, **neuron differentiation**)<sup>126</sup>. Two of our splicing factors are heterogeneous nuclear ribonucleoproteins (HNRNPs), as are three other prioritized genes (*HNRNPA1P3*, *HNRNPD*, *HNRNPUL2*)<sup>132,133</sup>. *HNRNPK* forms a complex with a number of other gene products, and its deletion in mice inhibits the proliferation of embryonic stem cells and neural progenitors<sup>134</sup>. *HNRNPK* may bring about its effects through multiple post-transcriptional mechanisms<sup>135</sup>; for instance, perturbations of its complex alter the expression of many genes, and one study found that *HNRNPK* and members of the ELAVL family (including the protein encoded by *ELAVL2*) are mutually antagonistic controllers of the switch from proliferation to neuronal differentiation through post-

transcriptional regulation of *CDKN1A* (also known as *P21*)<sup>136</sup>. This latter gene encodes an inhibitor of *CDK2* and *CDK4* in the cell cycle; it is also regulated by *PCBP4*<sup>137</sup>.

Several eukaryotic translation initiation factors are encoded by our prioritized genes (*EIF4A1*, *EIF4E*, *EIF4G1*, *EIF4G3*, *EIF4ENIF1*, *EIF5A*, *EIF5B*, *AGO2*). When the interaction between *EIF4G1* and *EIF4E* is disrupted, the latter factor loses much of its affinity for proliferation-promoting transcripts with a certain 5' motif, leading to the selective suppression of their translation<sup>138,139</sup>. The activity of the mTORC1 complex (which contains a subunit encoded by *RPTOR*) brings about the *EIF4G1*-*EIF4E* association and has been shown to amplify the proliferation of neural progenitors<sup>140–143</sup>. Under conditions of reduced mTORC1 activity and arrest of the cell cycle, some transcripts can still be connected to the translation machinery through interactions with a complex that includes the products of *AGO2*, *FXR1*, and *PARN*<sup>144</sup>. *AGO2* (formerly known as *EIF2C2*) can contribute to post-transcriptional regulation both in the cytoplasm and the nucleus; the gene is a high-ranking member of **nuclear import**, and it can be shuttled to the nucleus by interacting with the navigator protein encoded by *TNRC6A*<sup>145,146</sup>.

The placement of **mRNA splicing** and **regulation of nuclear mRNA splicing, via spliceosome** in **Supplementary Figure 22** suggests that splicing regulation exerts a relatively strong effect on brain development. However, alternative splicing—and indeed most forms of transcriptional regulation—occur at all times, and a number of splicing regulators (e.g., *RBFOX1*<sup>147,148</sup>) target the transcripts of genes expressed during synaptic function.

*Early Brain Development: Paracrine Influences on Cell Proliferation and Survival.* Using a different bioinformatics tool than we use, the most recent GWAS of intracranial volume found enrichment of several Reactome gene sets defined by the PI3K/AKT signaling pathway<sup>149</sup>. We find a similar pattern in our GWAS of *EduYears*: the cluster named after **GAB1 signalosome** contains the additional gene sets **PI3K events in ERBB2 signaling**, **PI3K events in ERBB4 signaling**, **PI3K/AKT activation**, **PIP3 activates AKT signaling**, and **signaling by ERBB4**. The clusters **downstream signal transduction** and **signaling by NGF** contain many of the other gene sets found to be significantly enriched in the GWAS of intracranial volume<sup>y</sup>. Although the PI3K/AKT

---

<sup>y</sup> Curiously, in contrast to the findings of the GWAS of intracranial volume, many gene sets in the DEPICT inventory referring to the generic cell cycle (e.g., **cell cycle** as opposed to **neural precursor cell proliferation**) are not significantly enriched in our study. On the other hand, these gene sets for the most part do not appear in our lists of negatively enriched sets; **Supplementary Table 27** contains **regulation of mitosis** and **negative regulation of mitosis**, and the negative results in **Supplementary Table 30** do not contain sets with the strings “cell cycle” or “mitosis” in their identifiers.

pathway seems to have been studied more intensively by cancer researchers than neuroscientists<sup>150–152</sup>, the literature nevertheless points to many connections between the action of this pathway and the ultimate size of the brain. The PI3K/AKT pathway may be active at many points in development, but we emphasize its likely influences on brain size because of its placement in **Supplementary Figure 22** (between the clusters **abnormal cerebral cortex morphology** and **regulation of nervous system development**) and its prominence in the just-cited GWAS of intracranial volume.

The PI3K/AKT signaling pathway is activated by the binding of particular paracrine factors to receptor tyrosine kinases, which include **IGF1R** (activated by the insulin-like growth factor **IGF1**) and **NTRK2** (activated by the brain-derived neurotrophic factor **BDNF**)<sup>153,154</sup>. The ligands activating these receptor tyrosine kinases are secreted by the *choroid plexus* (a network of blood vessels in the ventricle that produces the cerebrospinal fluid)<sup>155</sup>, other blood vessels innervating the brain<sup>156</sup>, and the neural progenitors themselves acting in autocrine fashion<sup>157</sup>.

Activated receptors recruit the adaptor protein GAB1, which in turn initiates the sequential recruitment of PI3K subunits. The PI3K enzyme has both a regulatory and a catalytic subunit, and genes encoding both subunits are located in our DEPICT-defined loci (**PIK3R1**, **PIK3R2**, **PIK3R3**, **PIK3C2B**, **PIK3C3**). PI3K enzymes fall into three classes. **PIK3R1** is the typical regulatory subunit employed in members of the well-studied class I; **PIK3C2B** is the catalytic subunit of class II members, which have been implicated in cell migration and survival; **PIK3C3** is the catalytic subunit of the sole class III member, which is involved in vesicular traffic, nutrient sensing, and MAPK signaling<sup>152</sup>.

Partially activated AKT (also called protein kinase B) suffices to activate mTORC1 by directly phosphorylating and inactivating an inhibitory complex composed of proteins encoded by **TSC1** and **TSC2**. (**GAB1 signalosome** and **TOR signaling cascade** are rather close in **Supplementary Figure 22**.) **AKT3** is encoded by a DEPICT-prioritized gene, as is the AKT-interacting protein **AKTIP**. Recall from our earlier discussion of post-transcriptional regulation that the activity of the mTORC1 complex ultimately increases the synthesis of proteins promoting proliferation. **TSC1/2** acts as a GTPase to convert the mTORC1 activator encoded by **RHEB** from its own GTP-bound state to its inactive GDP-bound state<sup>158</sup>. (Like ATP, GTP is nucleoside triphosphate whose hydrolysis provides energy to drive cellular reactions. The gene-set clusters **regulation of GTP catabolic process** and **GTPase regulator activity** are defined by this type of reaction.)

The kinase encoded by **PRKDC** (also known as **DNAP5**) can phosphorylate AKT at another position and thereby switch the kinase to its fully active form<sup>159</sup>. Fully active

AKT mediates numerous cellular functions, including proliferation through pathways other than the regulation of translation initiation via mTORC1. One target of fully active AKT is *MDM4* (also called *MDMX*), which AKT-mediated phosphorylation renders less liable to degradation<sup>160</sup>. *MDM4* binds to the anti-proliferative transcription factor TP53 (also called P53) and inhibits its activity through a variety of mechanisms: the blocking of transcriptional co-activators, removal from the nucleus, and exposure in the cytoplasm to degradative proteins<sup>161,162</sup>. Fully active AKT can also induce the sequestration of the cell-cycle inhibitors CDKN1A (also called P21) and *CDKN1B* (also called *P27*) from the nucleus; earlier we mentioned CDKN1A as a target of post-transcriptional regulation by *ELAVL2* and *PCBP4*.

Another effect of PI3K/AKT signaling on the targeted cell, somewhat distinct from the promotion of mitosis, is avoidance of *apoptosis* or *programmed cell death*. (This process defines the gene sets in the cluster **neuron apoptotic process**.) The pervasiveness of apoptosis implies that even moderate regulation of this process (e.g., sparing of neural progenitors) can have an impact on the ultimate size of the brain<sup>163</sup>. PI3K/AKT signaling is an important means by which paracrine factors (e.g., *IGF1*) can inhibit apoptosis. If phosphorylated by the fully active form of AKT, the FOXO transcription factors are sequestered away from the nucleus and thereby prevented from promoting the transcription of genes whose products inhibit the cell cycle (e.g., *CDKN1B*) and activate apoptosis<sup>153,164</sup>. (While *FOXO3* is a typical FOXO in this respect, *FOXO6* contains fewer phosphorylation sites and appears to fulfill other roles in brain development, including dendrite morphogenesis<sup>165,166</sup>.) AKT-mediated phosphorylation also inhibits CASP9, a protease that plays a key role in the cascade leading to the fragmentation of the DNA in the nucleus. *APAF1* encodes a component of a complex that cleaves the precursor of CASP9 and releases its mature form. AKT-mediated phosphorylation of *BAD* creates binding sites for the chaperone molecule *YWHAQ* (also known as *14-3-3*), which prevents *BAD* from performing its role in the mitochondrial release of cytochrome *c*—another component of the complex including *APAF1*. There appears to be extensive cross-talk between the PI3K/AKT and NFKB signaling pathways, such that each can upregulate the other<sup>167</sup>, and *HIVEP2* (formerly *MIBP1*) encodes a repressive transcription factor that inhibits genes involved in NFKB signaling<sup>168</sup>.

Once subject to the influence of paracrine factors, one way for a cell lineage to escape from this influence (even in the midst of ongoing ligand secretion) is for receptors such as *IGF1R* that are distributed non-uniformly on the membrane of the progenitor cell to be distributed asymmetrically to the two daughter cells as a result of cleavage along a plane that leaves more receptors on one side than the other<sup>169</sup>. (The disinherited daughter



is then freed to follow some other fate, such as leaving the cell cycle and becoming an astrocyte.) Within the cell itself, the phosphatase encoded by *PHLPP2* can dephosphorylate AKT and thus restore it to a less active form.

There is abundant evidence that perturbations of the PI3K/AKT pathway do indeed affect the ultimate number of developed neurons and their lamination<sup>170</sup>. Disruption of either *IGF1* or *IGF1R* lead to a retardation of brain growth in both mice and humans<sup>171</sup>. Deletion of *AKT3* in mice leads to a reduction of brain size<sup>172</sup>, and both germline and somatic mutations of *AKT3* in humans have been associated with microcephaly and agenesis of the corpus collosum accompanied by intellectual disability<sup>173</sup>, *megalencephaly* (enlargement of the brain)<sup>174</sup>, and *hemimegalencephaly* (enlargement of one cerebral hemisphere, accompanied by intellectual disability and epilepsy)<sup>175,176</sup>. *De novo* mutations of *PIK3R2* are thought to be the most frequent cause of a syndromic form of megalencephaly accompanied by various other neural and bodily malformations. *TSC1* and *TSC2* are named after tuberous sclerosis complex, a syndromic disorder caused by *de novo* mutations of the gene and characterized by hemimegalencephaly, abnormal neuronal migration, and intellectual disability; in mice, conditional knockout of *TSC1* in neural progenitor cells leads to an increase in mTORC1 signaling and brain size<sup>177</sup>. We earlier cited reports of mTORC1, the regulatory target of the TSC complex, being in turn a regulator of progenitor proliferation and brain size. Deletion of *AGAF1* in mice leads to several abnormalities that include brain overgrowth<sup>178,179</sup>. *FOXO3* was highlighted in the most recent GWAS of intracranial volume<sup>149</sup>; rs2022464, the sentinel SNP associated with intracranial volume at genome-wide significance, is concordantly associated with *EduYears* at a less stringent threshold ( $P = 4.6 \times 10^{-5}$ ). *De novo* mutations of *HIVEP2* have been implicated in intellectual disability occasionally accompanied by microcephaly<sup>180–182</sup>.

*Early Brain Development: Genesis and Exodus of Neurons.* Many of the genes involved in radial and tangential neuronal migration (the processes defining the exemplary gene set **telencephalon cell migration**) appear to be shared, but some are unique to one of these migratory modes. For example, *LHX6* encodes a transcription factor that regulates the expression of genes whose products are required by postmitotic interneurons undertaking migration and differentiation<sup>183,184</sup>.

The ligand encoded by *RELN* is perhaps the best-known guidance cue regulating neuronal migration<sup>185</sup>. *RELN* is a large glycoprotein secreted by *Cajal-Retzius cells* in the *marginal zone*, the topmost layer of the developing neocortex. The concentration of *RELN* declines from the marginal zone to the ventricle, and the receptors on the membrane of the newborn neuron encoded by *VLDLR* and *LRP8* (also known as *APOER2*) act as sensors of this gradient. When any of these genes are knocked out in

mice, the resulting phenotypes include ataxia, a reeling gait (hence the name of the ligand), and inverted lamination of the neocortex as a result of the inability of later-born neurons to climb past their earlier-born cousins. Unfortunately, whether **RELN** acts as a “stop” or “go” signal (this is possibly location dependent<sup>186</sup>) and how the disruption of its signaling leads to the observed abnormalities is still not well understood. What follows are some hints provided by several studies<sup>187,188</sup>.

The binding of **RELN** to its receptors induces phosphorylation of the adaptor protein **DAB1** by the kinases **SRC** and **FYN**. Phosphorylated **DAB1** can then go on to affect migration in any number of ways. For instance, it can interact with another adaptor protein, **CRK**, which in turns recruits **RAPGEF1**, the guanine nucleotide exchange factor for the small GTPase **RAP1A**. This GTPase appears to increase the membrane localization of the cell adhesion molecule encoded by **CDH2** (also called **NCAD**).

A *cell adhesion molecule* (CAM) is a surface protein that fastens the expressing cell to another cell. Cadherins are an important family of CAMs<sup>189</sup>, several of which are encoded by our prioritized genes (**CDH2**, **CDH4**, **CDH6**, **CDH7**, **CDH8**, **CDH9**, **CDH10**, **CDH12**, **CDH15**, **CDH18**, **CDH20**). The intercellular clasp formed by a given cadherin is formed by tokens of the cadherin extending from both cell membranes; the respective extracellular domains typically engage in *homophilic* binding of like extracellular domains at the region of contact. A migrating neuron is in constant contact with the basal process of a radial glial cell or whatever substrate it happens to be crawling on, and thus the forming and breaking of adhesive connections is an important process to regulate. A cell-cell junction formed by cadherins is known as an *adherens junction*, which has additional components on the cytoplasmic side called catenins (some of which are encoded by **CTNNA2**, **CTNNB1**, and **CTNND2**). Adherens junctions containing **CDH2** seal together apical progenitor cells into a skin-like barrier to the cerebrospinal fluid<sup>190</sup> and lead to a similar tight (but transient) packing of young neurons when they arrive at the **RELN**-dense layers of the developing cortex<sup>191</sup>. It is not yet clear how this temporary aggregation contributes to the lamination of the cortex.

Another branch of the **RELN/DAB1** pathway intersects with **PI3K/AKT**. Phosphorylated **DAB1** recruits **PIK3R1**; recall that dyslamination of the cortex is one of the phenotypes that can result from perturbations of the **PI3K/AKT** pathway<sup>192</sup>. The mechanism by which migration is affected seems to be the modulation of microtubule-binding proteins encoded by genes such as **MAPT**. Microtubules, as the largest discrete component of cytoskeleton, must be rearranged to facilitate whole-cell movement. **PI3K/AKT** activation also modulates the assembly and disassembly of actin filaments, the smaller elements of the cytoskeleton; one pathway terminus is the actin-depolymerizing protein encoded by **CFL1**. (Note that this gene and others whose



products work on the actin cytoskeleton are also involved in other processes involving fine cytoskeletal adjustments, such as the resizing of dendritic spines in synaptic plasticity.) Phosphorylated **DAB1** interacts directly with **PAFAH1B1** (also **LIS1**), another microtubule-associated protein. Mutations of **RELN** and **PAFAH1B1** can cause *lissencephaly*—a disorder characterized by defective neuronal migration, lack of brain gyri, and often intellectual disability<sup>193</sup>. **PAFAH1B1** is found at the *centrosome*, an organelle that serves as a hub of microtubule anchoring and participates in the regulation of nucleokinesis via centrosome-nucleus coupling. The microtubule-associated proteins encoded by **DCKLI** may play similar roles. **SUN1** encodes a nuclear-envelope protein that is also a part of the centrosome-nucleus coupling complex during nucleokinesis<sup>194</sup>.

Besides **RELN**, another well-studied guidance cue is **NRG1**. One of its receptors, **ERBB4**, is strongly expressed by tangentially migrating interneurons. The corridor from the GE to the dorsal cortex is lined by cells expressing high levels of membrane-bound **NRG1**, drawing the migrating interneurons along its length. **NRG1** secreted by neocortical cells then attracts the interneurons to the neighborhoods of their destinations.

The cyclin-dependent kinase encoded by **CDK5** also phosphorylates a number of different targets involved in neuronal migration. These include the actin regulator encoded by **PAKI**, which promotes the extension of the leading process. **CDK5**-knockout mice show an inversion of neocortical lamination similar to the **RELN** phenotype. The kinase phosphorylates **MAPT** and other microtubule-associated proteins, suggesting that **CDK5** is also a regulator of nucleokinesis.

Ephrins are a family of guidance molecules that bind to Eph receptor tyrosine kinases. Both ephrins and Ephs can fill the role of either ligand or receptor, depending on the context, and the interaction can result in a variety of cell movements including both attraction and repulsion. Ephrin-Eph interactions (involving specifically the products of **EPHB1**, **EPHB2**, **EFNA5**, **EFNB2**, and **EFNB3**) mediate the mutual repulsion that leads to the uniform distribution throughout the marginal zone of the **RELN**-secreting Cajal-Retzius cells<sup>195</sup>. Ephrins and Ephs also regulate tangential migration. For example, interneurons expressing **EPHA4** are repulsed from the cells making up the walls of their migration corridor by their expression of **EPHA5**, and this repulsion keeps them on their journey to the neocortex. Expression of **EPHA4** also repels interneurons from each other, ensuring that separate streams of migration stay segregated.

Transcriptional regulation is required to supply a neural progenitor cell with the gene products needed for proliferation, and it is required again to supply a newborn neuron with the products needed for its initial tasks of migration and adoption of a functional identity<sup>196–198</sup>. Diffusible paracrine factors may instruct certain neural

progenitors to begin transcribing the genes required for a dedicated postmitotic function even before neurogenesis. It is thus possible that the ventricular zone may be a mosaic of progenitor cells varying in the degree of multipotency<sup>199</sup>. Whatever their origin may be, gradients across the ventricular zone of **NR2F1** (also known as **COUPTF1**) and other transcription factors do appear to affect postmitotic neuronal identity. For example, conditional deletion of **NR2F1** in mice leads to an invasion by motor cortex of areas ceded by sensory cortex.

The transcription factor encoded by **SOX5** regulates the migration, differentiation, and axonal projections of neurons destined for the lower layers of the neocortex. Knockdown of this gene in mice leads to a number of aberrations, including overexpression of the transcription factors FEZF2 and **BCL11B** (also known as **CTIP2**), misrouting of corticothalamic axons to the hypothalamus, lack of projections to the pons and spinal cord, and a laminar inversion of the deep layers similar to the **RELN** phenotype. A revealing difference, however, is that late-born upper-layer neurons in **SOX5**-deficient mice migrate normally. This suggests that transcriptionally driven mechanisms of migration can be, after a certain point, cell autonomous—i.e., unaffected by aspects of the extracellular environment, such as the proper lamination of earlier-born neurons.

Another transcription factor that co-regulates multiple stages of the early neuronal career is encoded by **TBR1**. Perturbation in mice shows that **TBR1** has many functions in deep-layer neurons that are somewhat similar to those of **SOX5**. The abnormalities of neuronal migration observed in **TBR1**-deficient mice, however, are more complex in that the ectopic (“lost” or “trespassing”) cells are more deeply positioned in the frontal cortex and more superficially in the caudal cortex. **TBR1** may thus play a role in the specification of a rostral-caudal (anterior-posterior) as well as a layer-specific identity.

The laminar positioning and identity of upper-layer neurons are controlled by transcriptional programs that are distinct from those of their deeper-layer kin. One of the more important transcription factors is a chromatin modifier encoded by **SATB2**, whose regulated genes include **CUX1**, **CUX2**, **CDH10**, **RORB**, **AUTS2**, **UNC5C**, **EPHA4**, **TBR1**, and **BCL11B**<sup>200, z</sup> (**SATB2** is in turn an interaction partner of **MTA2**<sup>201</sup>, a component of the NuRD chromatin-remodeling complex discussed earlier.) When **SATB2** is perturbed in mice, early-born neurons migrate normally to the deep layers while later-born neurons are delayed in their arrival at the upper layers. The effect of **SATB2** on the deep-layer

---

<sup>z</sup> The CUX proteins are transcription factors highly expressed in progenitors giving birth to upper-layer neurons and in callosal projection neurons. **AUTS2** is an ASD susceptibility gene with strong memberships in **central nervous system neuron differentiation** and **regulation of neurogenesis**. **UNC5C** encodes a receptor of netrin, a guidance molecule directing neuronal migration and axon guidance.

transcription factor **BCL11B** is apparently repressive; in **SATB2**-deficient mice, **BCL11B** is ectopically transcribed in the upper layers. Consistent with this change in laminar identity, cortico-cortical axons are misrouted to subcortical regions, leading to the absence of the corpus callosum.

One way to summarize these findings (and others) is to think of the proteins encoded by **SOX5**, **TBR1**, **FZF2**, **BCL11B**, and **SATB2** as key nodes in a transcriptional network, often mutually repressing one another to sharpen the boundaries between functionally distinct neocortical layers. **TBR1** (expressed in the neurons of one deep layer) and **FEZF2** (expressed in the neurons of another deep layer) mutually repress one another; **SOX5** upregulates **TBR1** but represses **FEZF2**; **FEZF2** upregulates **BCL11B** but represses **SATB2** (expressed in neurons of upper layers), which returns the favor by suppressing **BCL11B**. Interestingly **SATB2** may begin to turn on the expression of **TBR1** in upper-layer neurons, particularly in the postnatal period, which suggests that refinements of neuronal identity continue even after the individual's birth.

The product of **POU3F2** (formerly **BRN2**) is also a transcription factor that affects gene expression across multiple stages (neurogenesis, migration, post-migratory differentiation)<sup>80,202</sup>. In contrast to **TBR1**, however, **POU3F2** activity occurs primarily in later-born neurons. Two important targets of **POU3F2** are **CDK5** and **DAB1**, and the fact that earlier-born neurons migrate normally in **POU3F2**-knockout mice reinforces the notion that neurons destined for different layers utilize distinct transcriptional programs to control the expression of common signaling molecules such as **DAB1**. Another target of **POU3F2** is **FOXP2**<sup>203</sup>, the so-called “language gene,” which encodes a transcription factor whose precise role in the brain has been somewhat elusive but that has been implicated in radial neuronal migration<sup>204,205</sup>. **FOXP1** is the closest homolog of **FOXP2**, and the two genes may participate in similar functions<sup>206,207</sup>. Another forkhead box transcription factor with multiple roles in brain development, including the regulation of neuronal migration, is encoded by **FOXG1**<sup>208,209</sup>. **NEUROD2** encodes another transcription factor expressed across several consecutive stages; its targets include **RELN**, **LRP8**, and **CUX1**<sup>210</sup>.

Although many subunits are degraded if unassociated with the **SMARCC1** component of the npBAF complex<sup>211</sup>, it appears that at least some can act independently as transcription factors. For instance, apart from its role in the BAF complex<sup>212</sup>, **BCL11A** regulates neuronal migration<sup>213</sup>. *De novo* mutations of **BCL11A** can result in intellectual disability accompanied by microcephaly, *pachygyria* (a disorder of neuronal migration similar to lissencephaly), and abnormalities of the corpus callosum<sup>214,215</sup>.

Intriguingly, note that many genes in the PI3K/AKT pathway (and others) are *oncogenes*—genes that in certain circumstances, most commonly a somatic mutation, become carcinogenic. This perhaps explains the adjacent positions of **GAB1** **signalosome** and **endometrial cancer** in **Supplementary Figure 22**. Cancer is a disease of dysregulated proliferation, evasion of apoptosis, and invasive migration, and one possibility may be that the mechanisms used to construct the brain in early life can go awry later to produce cancer in various tissues. This possibility is of course highly speculative but may be worthy of future investigation.

*Early Brain Development: Formation of Axons, Dendrites, and Synapses.* The targets of its axonal arbor are a crucial element of a post-migratory neuron's identity, and many of the mechanisms driving neuronal migration and axon guidance are shared. (The **STK11/STRADA/STK25/GOLGA2** pathway is an antagonist of **RELN/DAB1** signaling, and together they regulate the balance between the successive stages of migration and axon growth<sup>187</sup>.) Indeed, almost every gene known to influence axon guidance at the time of a recent review<sup>216</sup> has also been implicated in neuronal migration (especially tangential migration). The chief difference is obviously the absence of nucleokinesis in axon guidance, during which the soma is stationary. The description of axon guidance in Okbay et al.<sup>1</sup> suffices for our coarse-grained account, and thus we do not add to it here—except to note that the use of ROBO receptors by axonal growth cones to sense guidance molecules, represented in the previous paper by the significantly enriched gene set **signaling by ROBO receptor**, is now also represented by the individual genes **ROBO1** and **ROBO2**. **Signaling by ROBO receptor** and **axonogenesis** were distinct clusters in Okbay et al.'s results, but in the current results they are merged into the **axon guidance (Reactome)** cluster. Gene sets defined by axon growth are also present in the **neuron recognition** and **regulation of neuron projection development** clusters.

We now turn to how a neuron chooses one of its budding neurites to become its axon in the first place. The tendency of an *in vivo* neuron to select a process extending in a certain direction (e.g., toward the ventricle) to become its axon suggests that extrinsic cues bias the selection. In fact, the activation of the PI3K/AKT signaling pathway (possibly mediated by **BDNF/NTRK2**) appears to be a major mechanism of axon specification<sup>217,218</sup>. **SHTN1** (called **KIAA1598** in the DEPICT inventory) encodes a PI3K-interacting protein; its overexpression leads to the abnormal appearance of multiple axons, whereas its suppression inhibits axon formation. **SHTN1** is likely to be selectively transported to the nascent axon, therein regulating PI3K that has been activated by events at the plasma membrane. Accumulation of **SHTN1** in the growth cone accelerates the extension of the proto-axon, which in turn reduces diffusion of **SHTN1** back to the soma and leads to its further accumulation in the growth cone. Upon phosphorylation, the

ubiquitin ligase encoded by *SMURF1* singles out the growth-inhibiting GTPase encoded by *RHOA* for degradation rather than the axon-promoting factor encoded by *PARD6A*. (Actin-interacting proteins such as *RHOA* and *AKAP9* are involved in many different neuronal processes requiring the movement or reshaping of the neuron, including migration and synaptic plasticity.)

We turn now to the growth of a neuron's dendritic arbors<sup>219</sup>, the process defining the gene-set cluster **dendrite morphogenesis**. (The genes mentioned below also tend to be high-ranking members of sets in the cluster **regulation of neuron projection development**.) The branches of a dendrite appear to extend from its trunk with a high degree of stochasticity rather than being guided precisely to a distant destination. The branches tend to avoid each other and maximize the number of different axons that can be encountered. Self-avoidance is brought about by mutual repulsion of branches sprouting from the same neuron. Cell adhesion molecules (CAMs) embedded in the plasma membrane of a cell often serve as markers of personal identity for purposes of self/other discrimination. At least two sets of genes encoding CAMs are believed to mediate this type of discrimination in dendritic growth: the Down syndrome CAMs (*DSCAM*, *DSCAML1*) and the  $\gamma$  cluster of the protocadherins (*PCDHGC3*, *PCDHGC4*, *PCDHGC5*). In mice where these genes are null, sibling branches overlap and thus sample the local volume inefficiently. *DSCAM* and the  $\gamma$ -protocadherins share a number of features that underlie their utility in dendrite self-avoidance. Most importantly, they employ complex mechanisms to produce a large number of isoforms. In the case of *DSCAM*, the mechanism is alternative splicing. (*DSCAM* belongs to the immunoglobulin superfamily, which gets its name from the class of proteins that includes antibodies secreted by immune cells. Other immunoglobulin genes prioritized by our GWAS include *CHL1*, *CNTN3*, *CNTN4*, *CNTN5*, *SDK1*, *CADM1*, *CADM2*, and *CADM3*<sup>220,221</sup>.) The  $\gamma$ -protocadherins are arranged in a tandem array on chromosome 13, and each gene has a variable number of exons encoding extracellular, transmembrane, and proximal intracellular domains<sup>221</sup>. Tetramerization of  $\gamma$ -protocadherins leads to large number of isoforms in a manner similar to alternative splicing. Each neuron may well present a unique complement of *DSCAM* and  $\gamma$ -protocadherin isoforms serving as a barcode, although it is not yet known how homophilic contact between the extracellular domains protruding from different sibling branches is transduced into repulsion.

Transcriptional regulation is needed to supply appropriate gene products to the developing dendrites<sup>166,222</sup>. In fact, once the BAF complex has swapped out its proliferative components, its chief function appears to be the upregulation of genes whose products are employed in dendrite morphogenesis<sup>73</sup>, including *BDNF*, *NRN1* (also known as *CPG15*), *GAP43*, *RAP1A*, and *NGEF* (also *EPHEXIN1*). Knockdown of *SP4*,

which encodes a transcription factor, leads to the excessive branching of dendrites in cerebellar neurons. Terminating radial migration prematurely (by expressing a dominant-negative form of *CDH2* in neurons headed for the upper layers) causes a premature reduction in expression of the transcription factor encoded by *SOX11* and consequent increase in the number and length of dendrites<sup>223</sup>; this property of *SOX11* (which, as mentioned earlier, is a target of the BAF complex) again illustrates a transcriptional switch between earlier stages (proliferation, neurogenesis, migration) and later differentiation.

There must be some form of coordination or mutual influence to juxtapose the axonic side of the synapse (the bouton) and the dendritic side (the spine). In certain hippocampal pyramidal neurons, recognition molecules of the NGL/NTNG family produce a compartmentalization of the dendritic tree by ensuring the innervation of distinct portions of the tree by axons from correspondingly distinct sources. *NTGN1* and *NTGN2* encode ligands for receptors encoded by *LRRC4C* (also called *NGL1*) and *LRRC4* (also called *NGL2*), respectively. These latter two genes are members of the leucine-rich repeat family (*LRRC3B*, *LRRC4*, *LRRC4C*, *LRRC7*, *LRRTM4*, *LRRN2*, *LRRN3*, *LRRD1*, *LRFN2*, *LRFN4*, *LRFN5*, *LINGO1*, *LINGO2*, *SLITRK1*, *SLITRK5*, *FLRT1*, *LPHN3*), whose overarching function may be the exploitation of both family size and alternative splicing to create molecular fingerprints enabling precision wiring between neurons<sup>220,221</sup>. Incoming axons bearing *NTGN1* and *NTGN2* respectively somehow find their corresponding receptors (*LRRC4C*, *LRRC4*) and thereby innervate non-overlapping compartments on distal and proximal dendrites<sup>219</sup>. These leucine-rich repeats are synaptic CAMs that transmit signals to the dendritic interior upon transsynaptic binding, and thus their compartmental distribution may bestow varying information-processing properties along the length of a dendritic tree.

Neurexins are another class of synaptic CAMs, expressed on the presynaptic side (*NRXN1*, *NRXN2*, *NRXN3*). The vastly generative alternative splicing of a neurexin gene allows the extracellular domain to link up with various ligands expressed on the dendritic side, at least *in vitro* (*NLG1*, *DAG1*, *LRRTM4*, *LPHN3*)<sup>220,221,224</sup>. Determining the functions of neurexins and their ligands from knockout experiments in mice has been difficult because of frequent lethality, their possible redundancy in some contexts, and the variability of the neuronal phenotype. A recent study conditionally knocking out all three neurexin genes in different neuron types found evidence of context-dependent function; the phenotypes included disturbances in the distribution of synapses (Purkinje cells in the cerebellum), large decreases in synapse numbers (interneurons in a deep layer), and alterations of presynaptic spike-induced Ca<sup>2+</sup> transients (interneurons in the same layer but of a different type)<sup>224</sup>.



The neurotransmitter used at a mature synapse can initially function as a cue to the genesis of the synapse itself. The release of glutamate from an axon bouton can induce *de novo* growth of spines from a dendritic branch by opening NMDA-type glutamate receptors (*GRIN1*, *GRIN2A*, *GRIN2B*, *GRIN2D*, *GRIN3A*)<sup>225</sup>. As we discuss in more detail later, the opening of these receptors also triggers changes in signaling strength at the mature synapse. The internal pathways of synaptogenesis and synaptic plasticity appear to be different, however, in that the former does not rely on the kinase CaMKII. Perhaps surprisingly, the release of the neurotransmitter GABA can also induce the formation of dendritic spines<sup>226</sup>. The binding of GABA to its receptors typically inhibits the firing of the target neuron in the mature brain, but during development the extracellular environment is such that the current flow through opened GABA receptors can be excitatory. GABA-induced spine morphogenesis is mediated by depolarization of voltage-gated  $\text{Ca}^{2+}$  channels of the type encoded by *CACNA1D*, *CACNA1H*, and *CACNA1I*.  $\text{Ca}^{2+}$  influx can also induce the formation of inhibitory synapses by recruiting the anchor protein encoded by *GPHN*.

*Glycosylation* is a post-translational protein modification, like phosphorylation and ubiquitination. HNK-1, a glycan frequently attached to proteins in the brain, is composed of four domains, and all three of the enzyme types catalyzing the joining of one domain to the next are encoded by prioritized genes (*CHST10*, *B3GAT3*, *B4GALT2*)<sup>227</sup>. One substrate of HNK-1 is the AMPA-type glutamate receptor subunit *GRIA2*. Loss of HNK-1 on *GRIA2* results in weaker interaction of the subunit with *CDH2*, which may be the cause of the malformed dendritic spines lacking *GRIA2* that are also observed in the knockout mice. A behavioral correlate of these neuronal phenotypes is impaired memory. The enzyme-encoding genes are also high-ranking members of gene sets such as **neural precursor cell proliferation, regulation of neurogenesis**, and **telencephalon cell migration**, and evidence for the involvement of HNK-1 in the corresponding functions has emerged from studies of chicks and fish.

*ynaptic Communication and Plasticity: Transmembrane Transport of Ions and Other Solutes.* We now turn to our narrative summary of the neurophysiological mechanisms implicated by this GWAS of *EduYears* (**Supplementary Figure 8**). We begin with the genes that encode products involved in the transport of ions and other solutes across the neuronal membrane, a process corresponding to a number of gene sets (e.g., **ion transmembrane transport**) that reside mostly in the **gated channel activity** cluster. Gene products of this type are responsible for maintaining many of the background conditions required for information transfer in the nervous system, including the electrical and concentration gradients that enable current flow in and out of the neuron.

*ATP1A2*, *ATP1B1*, and *ATP1B3* encode subunits of the Na<sup>+</sup>/K<sup>+</sup> ATP-driven exchanger chiefly responsible for the much higher concentration of Na<sup>+</sup> in the extracellular fluid bathing the neuron than in the neuronal interior (and, conversely, much lower concentration of K<sup>+</sup>).

The superfamily of genes encoding solute carriers is extremely large and diverse, and many members are prioritized in our GWAS (*SLC2A13*, *SLC4A2*, *SLC4A10*, *SLC6A9*, *SLC6A10*, *SLC6A15*, *SLC6A17*, *SLC8A3*, *SLC9A5*, *SLC12A5*, *SLC22A23*, *SLC24A1*, *SLC24A2*, *SLC24A3*, *SLC25A12*, *SLC26A10*, *SLC29A4*, *SLC35E2*, *SLC35E4*, *SLC35F4*, *SLC38A1*, *SLC38A2*, *SLC39A10*, *SLC45A1*). Many of these encode transporters that also contribute to the ion concentrations found on the two sides of the neuronal (or organelle) membrane, although usually by relying on existing concentration gradients to drive the change (e.g., gradients set up by ATP-driven pumps). *SLC8A3* (formerly *NCX3*) encodes a Na<sup>+</sup>/Ca<sup>2+</sup> exchanger that typically extrudes one Ca<sup>2+</sup> ion for every three Na<sup>+</sup> ions taken in. *SLC24A1*, *SLC24A2*, and *SLC24A3* (also called *NCKX1*, *NCKX2*, and *NCKX3*, respectively) encode potassium-dependent Na<sup>+</sup>/Ca<sup>2+</sup> exchangers, which typically extrude one Ca<sup>2+</sup> ion and one K<sup>+</sup> ion for every four Na<sup>+</sup> ions taken in<sup>228,229</sup>. These exchangers operate in parallel with the ATP-driven Ca<sup>2+</sup> pump, an isoform of which is encoded by *ATP2B2*.

Cytosolic increases in Ca<sup>2+</sup> can also be cleared by sequestering the ions in intracellular organelles. There is a large potential difference across the mitochondrial inner membrane, as a consequence of the hydrogen ion (H<sup>+</sup>) concentration gradient generated by oxidative phosphorylation. This gradient can be used to transport Ca<sup>2+</sup> into the mitochondrion through channels encoded by *MCU*. Without necessarily entering the mitochondrion, Ca<sup>2+</sup> can also stimulate the synthesis of ATP by binding to a glutamate-aspartate exchanger in the inner mitochondrial membrane encoded by *SLC25A12*<sup>230</sup>.

*SLC12A5* (also known as *KCC2*) encodes the K<sup>+</sup>/Cl<sup>-</sup> cotransporter that is chiefly responsible for the hyperpolarizing effect of Cl<sup>-</sup> permeability in mature neurons, and its product extrudes one Cl<sup>-</sup> ion outward together with one K<sup>+</sup> ion<sup>231</sup>.

Active transport must also be used to regulate the intracellular pH because otherwise the neuronal interior would be rather acidic. Such regulation is probably obligatory because of the dependence of many parameters, such as the opening probability of ion channels, on intracellular pH. *SLC9A5* (formerly called *NHE5*) encodes a Na<sup>+</sup>/H<sup>+</sup> exchanger that alkalinizes the cell by extruding H<sup>+</sup>. It may be that the action of this particular exchanger is not a “background” activity, for it is recruited to dendritic spines by activation of NMDA-type glutamate receptors and plays a role in spine growth<sup>232</sup>.



Some of the carriers appear to transport neurotransmitters or their precursors into neurons. *SLC6A15* and *SLC6A17* are amino acid transporters expressed in the brain with a high affinity for leucine, which is used in the production of glutamate<sup>233,234</sup>. *SLC29A4* (formerly *PMAT*) encodes a reuptake transporter that clears the neurotransmitters serotonin and dopamine from the synaptic cleft, thereby terminating whatever signals they may be conveying<sup>235</sup>.

*SLC45A1* encodes a cerebral glucose transporter, and *de novo* mutations of the gene have recently been implicated in intellectual disability accompanied by epilepsy<sup>236</sup>. This finding draws attention to the possibly understudied role of metabolism and energetics in cognition<sup>237</sup>.

Whether synthesized in the cytoplasm or recycled from the synaptic cleft, neurotransmitter must be transported into synaptic vesicles. *ATP6V0A1*, *ATP6V0D1*, and *ATP6VID* encode transmembrane domain subunits of vacuolar ATPase; this complex actively transports H<sup>+</sup> into the vesicle. Because the movement of neurotransmitter into the vesicle is coupled to the movement of H<sup>+</sup> along its electrochemical gradient, the acidification of the vesicular interior ensures the filling of the vesicle with neurotransmitter<sup>238</sup>.

*Synaptic Communication and Plasticity: Release of Neurotransmitter into the Synaptic Cleft.* Given the properly calibrated extracellular and intracellular concentrations of the relevant ions and the presence of the required receptors, channels, and other gene products in its dendrites and axon, a cortical neuron is capable of carrying out its fundamental tasks of signal processing and transmission. We will begin our account of neuronal communication with the genes expressed in the axon bouton<sup>239</sup>. The genes mentioned here tend to be high-ranking members of gene sets in the clusters **voltage-gated calcium channel activity, associative learning, cytoplasmic vesicle membrane, regulation of synaptic transmission, serotonin neurotransmitter release cycle, regulation of neurotransmitter levels, and synapse part**.

*Exocytosis* is the fusion of one membrane (e.g., the vesicle membrane) with a second membrane (e.g., the plasma membrane of the cell). Exocytosis in the axon bouton occurs at *active zones* consisting of a dense collection of scaffold proteins connecting the cytoskeleton to the plasma membrane. These proteins include PCLO (piccolo), *BSN* (bassoon), and many others<sup>240</sup>. The RAB family of GTP-binding proteins regulates the budding and fusion of all membranous vesicles, including synaptic vesicles<sup>241</sup>. For now we are concerned with the RAB3 genes and *RAB3GAP2*, whose products participate specifically in neurotransmitter release<sup>242–244</sup>.

The RAB3 proteins (**RAB3A**, **RAB3B**, **RAB3C**, RAB3D) drive a GTP-dependent cycle. When a RAB3 is bound to GTP, it spends its time anchored to the membrane of a synaptic vesicle. The RAB3 is part of a larger complex, one slot of which can be filled by a UNC13 protein (e.g., **UNC13C**) and the remaining slot by **RIMS1** or **RIMS2**. (**RIMS3** is also strongly expressed in the brain, but its product lacks some of the domains present in the lower-numbered family members<sup>240</sup>.) The RIMS protein anchors the complex—and, thus, the synaptic vesicle—to a voltage-gated  $\text{Ca}^{2+}$  channel spanning the plasma membrane. The GTPase-accelerating RAB3GAP complex (a subunit of which is encoded by **RAB3GAP2**) stimulates RAB3's intrinsic GTP-hydrolysis activity so that its bound GTP is converted to GDP, and subsequently RAB3 dissociates from the vesicle membrane to begin the phase of its cycle spent as a soluble protein in the cytoplasm of the axon bouton. The suppression of this disassociation appears to inhibit the priming of exocytosis by keeping RAB3 anchored to the vesicle<sup>244</sup>. Under normal conditions the RAB3-bound GDP is exchanged for GTP in the cytoplasm, allowing the RAB3 to start another round of the cycle by associating with another vesicle that has not yet docked at the plasma membrane.

A docked and primed synaptic vesicle will fuse with the plasma membrane upon the influx of  $\text{Ca}^{2+}$  through a voltage-gated channel. A channel complex consists of the main pore-forming subunit  $\alpha_1$ , possibly in combination with auxiliary subunits ( $\alpha_2\delta$ ,  $\beta$ ). All of these subunit types are represented in our prioritized genes:  $\alpha_1$  (**CACNA1A**, **CACNA1B**, **CACNA1C**, **CACNA1D**, **CACNA1E**, **CACNA1H**, **CACNA1I**),  $\alpha_2\delta$  (**CACNA2D1**, **CACNA2D2**, **CACNA2D3**), and  $\beta$  (**CACNB1**, **CACNB2**, **CACNB3**, **CACNB4**).

**CACNA1A** and **CACNA1B** encode  $\alpha_1$  subunits often called **Cav2.1** and **Cav2.2** respectively, and most synapses in the central nervous system rely on channels containing these subunits for mediating rapid vesicular release from the axon bouton. In contrast, complexes containing the pores encoded by **CACNA1C**, **CACNA1D**, **CACNA1H**, and **CACNA1I**—**Cav1.2**, **Cav1.3**, **Cav3.2**, and **Cav3.3**, respectively—are mostly found on dendrites.

**CACNA2D1**, **CACNA2D2**, and **CACNA2D3** encode  $\alpha_2\delta$  auxiliary subunits that are widely expressed in the brain and form channel complexes with  $\alpha_1$  subunits in the Cav1 and Cav2 subfamilies<sup>245</sup>. The  $\alpha_2\delta$  subunits generally increase the expression of the  $\alpha_1$  subunits and can also increase their rate of inactivation following a voltage-induced opening.

The  $\beta$  subunit **CACNB3** is the predominant but not exclusive complex partner of **Cav2.2**, while **CACNB4** is the predominant partner of **Cav2.1**<sup>246</sup>. The  $\beta$  subunits also influence channel expression but have a much greater effect on physiological properties

of the channel complex. The presence of a  $\beta$  subunit leads in most cases to a larger current influx across the range of voltage where action potentials are likely and accelerated rates of both activation and inactivation.

Once an invading action potential has opened a voltage-gated  $\text{Ca}^{2+}$  channel complex, the influx of  $\text{Ca}^{2+}$  brings about the fusion of the vesicle with the plasma membrane by changing the conformation of the SNARE complex, a multi-protein structure that will also become attached to the docked vesicle<sup>240,247</sup>. (The gene sets **SNARE complex**, **SNARE binding**, and **proteolytic cleavage of SNARE complex proteins** are members of the cluster **ENSG00000179036 PPI subnetwork**.) Many components of the complex and associated proteins are encoded by prioritized genes (*VAMP1*, *VAMP2*, *VAMP3*, *STX1B*, *SNAP25*, *STXBP1*, *STXBP6*, *CPLX1*, *SYT1*, *SYT6*, *SYT7*, *SYT11*, *SYT12*, *SYT17*, *NSF*). The VAMP genes encode synaptobrevins, the component of the SNARE complex providing the attachment to the synaptic vesicle; most of its amino acids, however, protrude into the cytoplasm. Syntaxin, the product of *STX1B*, has a similar structure but is attached to the plasma membrane. (The gene set **syntaxin-1 binding** is a member of the cluster named after **serotonin neurotransmitter release cycle**.) *SNAP25* is also anchored to the plasma membrane by palmitoyl chains.

Prior to the docking of the vesicle, the SNARE complex is not assembled. Instead each of the components is bound to other proteins. For example, a syntaxin-binding protein, such as *STXBP1* (also known as *MUNC18-1*) or *STXBP6*<sup>248,249</sup>, can prevent syntaxin from being a part of the SNARE complex. (The regulatory function of *STXBP1* is not purely negative, as it is essential for neurotransmitter release<sup>250</sup>.) After the docking of the synaptic vesicle within close proximity of a voltage-gated  $\text{Ca}^{2+}$  channel, the SNARE components become unbound from their alternative partners and form the complex. When the three main components (synaptobrevin, syntaxin, *SNAP25*) first combine, the complex resides in the *trans* state, which is energetically very unfavorable. It is the subsequent collapse of the SNARE complex into the energetically more favorable *cis* state that fuses the vesicle with the plasma membrane. Given a baseline local concentration of  $\text{Ca}^{2+}$ , the SNARE complex is kept in the *trans* state by *CPLX1*; this protein binds to the groove between syntaxin and synaptobrevin and prevents the complex from collapsing into the energetically preferred *cis* state.

The proteins of the SNARE complex are themselves insensitive to  $\text{Ca}^{2+}$ . The protein that detects the spike-induced  $\text{Ca}^{2+}$  influx is synaptotagmin, which is associated with the vesicle membrane but has two domains protruding into the cytoplasm that bind  $\text{Ca}^{2+}$ . When this binding occurs, the synaptotagmin is thought to displace *CPLX1* from the SNARE complex, triggering its collapse into the *cis* state and thus the ejection of the vesicle's neurotransmitter contents into the synaptic cleft. *SYT1* is the predominant  $\text{Ca}^{2+}$ -

sensing synaptotagmin used to trigger vesicular release in tight synchrony with action potentials<sup>251,252</sup>. Neurons sometimes release transmitter spontaneously or in loose synchrony with the action potential<sup>253</sup>, and **SYT7** may be a slower-acting  $\text{Ca}^{2+}$  sensor mediating asynchronous release<sup>254</sup>. **SYT7** also appears to be required for *facilitation*, a form of synaptic strengthening in which each subsequent action potential in a closely spaced series evokes greater neurotransmitter release<sup>255</sup>. **STY11** and **SYT12** do not bind  $\text{Ca}^{2+}$  and therefore may regulate a form of vesicle exocytosis that does not depend on the influx of this ion<sup>255</sup>.

After exocytosis, the SNARE complex must be disassembled in order to reinitiate the release cycle. The set of proteins responsible for the disassembly include **NSF** and soluble **NSF** accessory proteins (SNAPs)<sup>256,257</sup>. (Note that these SNAPs are unrelated to **SNAP25**.) SNAPs wrap around the elongated SNARE complex, and several instances of **NSF** assemble at one end of the resulting SNAP-SNARE complex. ATP hydrolysis unsprings the **NSF** portion of the complex and in this way pries the SNARE components apart from each other.

Once a vesicle has released its contents by fusing with the plasma membrane, retrieval mechanisms fetch the membrane lipids of the vesicle and its associated proteins that have become embedded in the plasma membrane. The retrieved membrane materials can be recycled to form new vesicles, which in turn can be refilled with neurotransmitter. An important retrieval mechanism is *clathrin-mediated endocytosis*, in which clathrin proteins coat the targeted material in a chicken-wire-like cage, which is then pinched off from the rest of the plasma membrane<sup>258</sup>.

The first step in the synthesis of the clathrin cage is recruitment to the plasma membrane of clathrin and adaptor proteins, which bind to the remnants of the recently fused vesicle. This recruitment is accomplished by the phosphatidylinositol-binding protein encoded by **PICALM** (formerly **CALM**, which is now the symbol of an unrelated gene family). **AP2B1** encodes a subunit of an adaptor protein that targets synaptotagmins. Binding of adaptor proteins to their targets requires the phospholipid  $\text{PIP}_2$ , which is found only in the plasma membrane. This ensures that clathrin coats do not enclose internal organelles such as the free vesicles themselves<sup>259</sup>.

The binding of adaptor proteins not only recruits clathrin to the membrane but also another set of proteins that can include **EPN2** and **SH3GL1**, which warp the targeted portion of the plasma membrane into a new spherical vesicle around which the clathrin cage continues to assemble. Finally, the dynamins encoded by **DNMI** and **DNM3** detach the newly caged vesicle from its place of origin in the plasma membrane by forming and cinching a collar around the neck of the vesicle.

*Synaptic Communication and Plasticity: Ionotropic Neurotransmission.* The glutamate receptor of the AMPA type is employed at a majority of excitatory synapses in the brain, and it bears the chief responsibility for driving the immediate depolarization of the dendrite in response to incoming signals<sup>260</sup>. The importance of fast excitatory neurotransmission is supported by the exemplary gene set **extracellular-glutamate-gated ion channel activity**.

Most AMPA-type glutamate receptors in the adult brain appear to be heterotetramers assembled from four possible subunits (**GRIA1**, **GRIA2**, GRIA3, **GRIA4**)<sup>261</sup>, usually consisting of **GRIA1** and **GRIA2**. There is some variability in the time course of the potential change induced by current flow through an opened AMPA-type glutamate receptor, depending on subunit composition, alternative splicing, and the presence of transmembrane AMPA-receptor regulatory proteins (TARPs). The depolarizing current flowing through open AMPA-type glutamate receptors is carried mostly by Na<sup>+</sup>, but receptors composed of **GRIA1**, GRIA3, and **GRIA4** show some permeability to Ca<sup>2+</sup> as well. **GRIA2**-containing receptors are typically not permeable to Ca<sup>2+</sup>.

Our prioritized genes include several that encode TARPs (**CACNG3**, **CACNG7**, **CNIH2**, **SYNDIG1**). **CACNG3** and **CACNG7** belong to a family of TARPs that are evolutionarily related to the  $\gamma$  subunit of the voltage-gated calcium channel complex. It can bind to AMPA-type glutamate receptors and has been implicated in a variety of functions: acting as a chaperone to traffic receptors to the dendritic membrane, increasing the mean channel conductance, and reducing the rates of deactivation and desensitization<sup>262,263</sup>. **CNIH2** also has been found to increase mean channel conductance, but its important effects on other aspects of receptor physiology appear to be quite variable.

The NMDA-type glutamate receptors take their name from the substance that was first shown to bind to them, the glutamate analog NMDA. Several genes encoding subunits of these receptors are among our prioritized genes (**GRIN1**, **GRIN2A**, **GRIN2B**, **GRIN2D**, **GRIN3A**). The GRIN2 subunits have a binding sites for glutamate, where the **GRIN1** and GRIN3 subunits bind glycine (or D-serine)<sup>264</sup>. The receptor as a whole is a heterotetramer, most commonly composed of two glycine-binding **GRIN1** subunits and two glutamate-binding GRIN2 subunits.

Subunits of the typically inhibitory GABA<sub>A</sub> receptor are grouped into subfamilies known respectively as  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\Theta$ ,  $\pi$ , and  $\rho$ ,<sup>265</sup> the first three and last of which are represented in our DEPICT-prioritized genes (**GABRA4**, **GABRA6**, **GABRB1**, **GABRB2**, **GABRB3**, **GABRG1**, **GABRG3**, **GABRR1**). A complete receptor is composed of five subunits; the **GABRB3** homopentamer was recently used in the first study to obtain an X-

ray structure of a GABA<sub>A</sub> receptor<sup>266</sup>, but a heteromeric arrangement such as GABRG2-GABRB2-GABRA1-GABRB2-GABRA1 (counterclockwise around the pore as viewed from the outside of the neuron) seems to be more typical. It happens that GABRA4, GABRB2, and GABRB3 all take part in pentamers with evidence of abundant use in the brain<sup>51,267</sup>. Some subunits of GABA<sub>A</sub> receptors contain regulatory sites for phosphorylation and domains that interact with trafficking proteins such as the one encoded by GPHN<sup>268</sup>. The products of GPHN and IQSEC3 interact to increase the density of inhibitory synapses<sup>269</sup>.

5-HT<sub>3</sub>, the one receptor of serotonin that is a ligand-gated ion channel, is a pentameric structural cousin of GABA<sub>A</sub>.<sup>270</sup> Two subunits that are found in both rodents and humans are encoded by HTR3A and HTR3B. 5-HT<sub>3</sub> is expressed most strongly in the brainstem, especially in areas involved in the vomiting reflex, and to a lesser degree in forebrain areas such as the amygdala and hippocampus<sup>271</sup>. In the hippocampus 5-HT<sub>3</sub> is found in the dendrites of GABAergic interneurons<sup>272</sup>. Homomeric 5-HT<sub>3A</sub> mediates rapidly activating and desensitizing inward currents, which are carried primarily by Na<sup>+</sup> and K<sup>+</sup>. 5-HT<sub>3B</sub> subunits do not form homomeric channels on their own, but their incorporation into heteromeric subunits with 5-HT<sub>3A</sub> leads to a complex with distinctive properties, including reduced permeability to Ca<sup>2+</sup> and greater single-channel conductance.

*Synaptic Communication and Plasticity: Na<sup>+</sup> and K<sup>+</sup> Channels.* Many of the genes that we now discuss are high-ranking members of gene sets in the cluster named after **gated channel activity**, including **voltage-gated sodium channel complex** and **potassium channel complex**.

The depolarizing current flowing into opened ionotropic receptors (mostly by glutamate) will trigger an output spike in the receiving neuron if the depolarization suffices (after taking into account any inhibition) to set off a chain reaction of opening voltage-gated Na<sup>+</sup> channels. The  $\alpha$  subunits of all five Na<sup>+</sup> channels that are strongly expressed in the central nervous system are encoded by our prioritized genes (SCN1A, SCN2A, SCN3A, SCN5A, SCN8A). The  $\alpha$  subunits of these respective channels (Nav1.1, Nav1.2, Nav1.3, Nav1.5, Nav1.6) contain the ion-conducting pore, the selectivity filter that prefers Na<sup>+</sup> over other ions, and the elements responsible for activation (voltage sensing) and inactivation<sup>273</sup>. These channels particularly enrich the unmyelinated *axon initial segment* (AIS) and the nodes of Ranvier<sup>274,275</sup>.<sup>aa</sup> This is consistent with the critical

---

<sup>aa</sup> More specifically, Nav1.1 enriches the AIS of interneurons in the hippocampus and cerebellum. Even on the patch of plasma membrane bounding the AIS of cortical pyramidal neurons, Nav1.2 and Nav1.6 have



role of voltage-gated Na<sup>+</sup> channels in the initiation and propagation of action potentials. Nav1.6 is also found in the plasma membrane bounding the soma and the roots of the dendritic trees, although their density at these locations is smaller than at the AIS<sup>276,277</sup>. The density of Nav1.6 along the apical dendrite further declines with distance from the soma, and the channel appears to be wholly absent from dendritic spines.

There are many different types of K<sup>+</sup> channels, each repolarizing the neuronal plasma membrane after an action potential in a distinct way or carrying out some entirely different function. Our prioritized genes include more than twenty that encode subunits of K<sup>+</sup> channels, including voltage-gated channels (*KCNC2*, *KCNC4*, *KCND3*, *KCNH1*, *KCNH2*, *KCNH3*, *KCNH5*, *KCNH6*, *KCNH7*, *KCNQ5*), ancillary proteins (*KCNAB2*, *KCNAB3*, *KCNIP1*, *KCNIP2*, *KCNIP4*, *DPP6*, *DPP10*, *KCNE2*), modifiers that do not form homomeric channels on their own (*KCNF1*, *KCNIG1*, *KCNS1*), ion-activated channels (*KCNMA1*, *KCNMB4*, *KCNN2*, *KCNT1*), and inward rectifiers (*KCNJ3*, *KCNJ4*, *KCNJ6*). More work needs to be done to characterize the regional and subcellular distributions of these channels. Below, we attempt to convey a consensus finding whenever this seems to be justified<sup>278–282</sup>.

Four subfamilies of voltage-gated K<sup>+</sup> channels have been recognized: *Shaker* (Kv1), *Shab* (Kv2), *Shaw* (Kv3), and *Shal* (Kv4). *KCNAB2* and *KCNAB3* encode auxiliary  $\beta$  subunits that forms channel complexes, for the most part, with *Shaker*  $\alpha$  subunits<sup>283,284</sup>. The precise pore-forming partner can depend on alternative splicing. *In vitro* studies have shown that the  $\beta$  subunit can affect current density and shift channel activation to more negative voltages, but its most striking effect in these studies is the addition of rapid inactivation where otherwise the channel does not inactivate at all. This addition can have profound consequences on the information-processing properties of the neuron (as discussed further below).

*KCNF1*, *KCNIG1*, and *KCNS1* respectively encode Kv5.1, Kv6.1, and Kv9.1, which cannot conduct current on their own but modulate the properties of *Shab* channels. These channels are found on the soma and dendrites of neurons throughout the brain. They are classified as voltage-dependent *delayed rectifiers*—so called because of slow activation upon membrane depolarization, in comparison to voltage-gated Na<sup>+</sup> channels. (This lag ensures that Na<sup>+</sup> influx precedes K<sup>+</sup> efflux, allowing an action potential to fire before repolarization sets in.)

---

somewhat different distributions; Nav1.2 preferentially enriches the proximal portion of the segment closer to the soma, whereas Nav1.6 enriches the distal portion bordering the first myelin sheath<sup>360</sup>.

Both *KCNC2* and *KCNC4* encode pore-forming  $\alpha$  subunits of *Shaw* channels (*Kv3.2*, *Kv3.4*). In contrast to the delayed rectifier *Kv3.2*, *Kv3.4* is a conduit for rapidly inactivating *A current*, and it has been reported to be present on the axons of projection neurons. Whereas delayed rectifiers are mostly responsible for the falling phase of the action potential, A-current channels like homomeric *Kv3.4* regulate the duration of the interval between spikes in a neuron that fires repeatedly. A wealth of empirical research has shown that delayed rectifiers (e.g., *Kv3.2*) and conductors of A current (e.g., *Kv3.4*) can fairly be considered extremes of a continuum<sup>279</sup>. By expressing different combinations of voltage-gated K<sup>+</sup> channels, neurons can generate a spectrum of delay times and periods governing its repetitive spiking in response to sustained input.

*KCND3* encodes the pore-forming  $\alpha$  subunit of *Kv4.3*, a *Shal* channel that admits A current into the soma and dendrites of interneurons in the neocortex and hippocampus. *Shal* channel complexes tend to contain interacting proteins of the KCNIP and DPP families. Co-expression of *KCNIP1*, *KCNIP2*, or *KCNIP4* with *KCND3* results in a channel complex that inactivates more slowly and recovers from inactivation more rapidly than the  $\alpha$  subunit expressed alone<sup>284</sup>. In addition, *KCNIP1* increases the density of the current flowing through *Kv4.3*, probably by increasing the trafficking of the channel to the plasma membrane. One study has found that deletion of one KCNIP tends to increase expression of others in a compensatory fashion, but that simultaneous knockdown of *KCNIP2*, *KCNIP3*, and *KCNIP4* does result in a reduction of A current<sup>285</sup>. The incorporation of either *DPP6* or *DPP10* into the complex, which can occur separately or together with a KCNIP, also increases current density and the rate of recovery from inactivation; the rate of inactivation, in contrast, decreases.

*KCNMA1* encodes the pore-forming  $\alpha$  subunit of *KCa1.1*, which is notable for being gated by Ca<sup>2+</sup> as well as by voltage. The probability of channel opening is low at negative membrane potentials but reaches nearly 100 percent upon both an increase in potential and the addition of Ca<sup>2+</sup>. The  $\alpha$  subunit resembles the *Shaker* channel in structure; one prominent difference is the attachment of a large structure on the cytoplasmic side with a Ca<sup>2+</sup>-binding site<sup>286–288</sup>. *KCa1.1* channels are positioned in either close proximity or physical contact with voltage-gated Ca<sup>2+</sup> channels (e.g., *Cav2.1*), and in fact this may be the only location where the Ca<sup>2+</sup> concentration can become elevated enough to trigger *KCa1.1* opening. When such elevation occurs in an axon bouton as a result of an action potential, the subsequent hyperpolarization produced by the efflux of K<sup>+</sup> accelerates the closing of the voltage-gated Ca<sup>2+</sup> channels and may act to limit the release of neurotransmitter.



*KCNN2* encodes a subunit of *KCa2.2*, another  $\text{Ca}^{2+}$ -gated  $\text{K}^+$  channel. Compared to *KCa1.1*, *KCa2.2* channels have a smaller conductance, weaker voltage dependence, and greater  $\text{Ca}^{2+}$  sensitivity<sup>289</sup>. *KCa2.2* channels play a role in a variety of processes other than hypolarization following an action potential; in many projection neurons, they reside in the postsynaptic membrane and influence synaptic plasticity.

*Inward rectifiers* are defined by the property of passing  $\text{K}^+$  current more readily in the inward direction regardless of the gradient. These channels are critical to setting the resting potential, permitting the action potential to proceed, and returning the membrane to rest. This basic housekeeping role is fulfilled by so-called classic inward rectifiers, such as *Kir2.3*; this channel is encoded by *KCNJ4*. We defer discussion of the *Kir3* family until we discuss metabotropic transmission.

Our prioritized genes encode all subunits of hyperpolarization-activated cyclic nucleotide-gated (HCN) channels (*HCN1*, *HCN2*, *HCN3*, *HCN4*). One function of HCN channels is the stabilization of the resting potential: hyperpolarization activates the channel and leads to depolarization, while depolarization closes the channel and cuts off its contribution. Consistent with this role, *HCN1* is found on the soma and dendrites of pyramidal neurons in the neocortex and hippocampus; in fact, its density increases with distance from the soma<sup>290</sup>. This distribution may have the effect, in this particular type of neuron, of normalizing excitatory input so that its contribution is less dependent on distance from the AIS. Another possible function is to generate rhythmic bursts of action potentials. Upon depolarization, the activation of channels such as *Cav3.2* may lead to dendritically initiated spikes. The overshoot of the repolarization following the action potential opens HCN channels, depolarizing the neuron again and leading to a fresh spike.

*Synaptic Communication and Plasticity: Metabotropic Neurotransmission.* Genes overlapping our DEPICT-defined loci encode subunits of metabotropic receptors activated by several distinct neurotransmitters: glutamate (*GRM1*, *GRM2*, *GRM3*, *GRM5*, *GRM7*, *GRM8*), serotonin (*HTR1A*, *HTR1B*, *HTR2A*), dopamine (*DRD1*, *DRD2*), and acetylcholine (*CHRM1*, *CHRM3*, *CHRM4*). What all of these receptor types have in common is the transduction of the signal conveyed by their ligand, via a *G protein*, into further changes at the membrane or in the depths of the neuronal interior.

A heterotrimeric G protein is composed of  $G\alpha$ ,  $G\beta$ , and  $G\gamma$  subunits; all three subunit types have representatives encoded by DEPICT-prioritized genes (*GNAI1*, *GNAI2*, *GNAI3*, *GNAT1*, *GNB1*, *GNG3*, *GNG7*)<sup>291</sup>. G protein diffuses freely in the plasma membrane, but binding to the appropriate activated metabotropic receptor

catalyzes the exchange of GDP for GTP at a guanine nucleotide-binding site on the  $G\alpha$  subunit. The subunits of the G protein then disassociate into  $G\alpha\cdot\text{GTP}$  and  $G\beta\gamma$ , both of which can interact with effectors that bring about some change in the information-processing properties of the neuron. One way the signal from the ligand can be terminated is the hydrolysis of the  $G\alpha$ -bound GTP and reassembly of the inactive  $G\alpha\beta\gamma$  trimer. The rate at which this hydrolysis occurs can be greatly increased by regulator of G-protein signaling (RGS) proteins, some of which are encoded by DEPICT-prioritized genes (*RGS6*, *RGS7*, *RGS12*, *RGS17*)<sup>292</sup>. *RGS6* and *RGS7* are members of an RGS family called R7, and their proteins may be anchored to the plasma membrane by the binding protein *RGS7BP*<sup>293</sup>.

The consequences of  $G\alpha\cdot\text{GTP}$  and  $G\beta\gamma$  being transiently separate depend on a variety of factors but touch on every aspect of neuronal function. Here we only mention a few illustrative examples. In some cases  $G\alpha\cdot\text{GTP}$  and  $G\beta\gamma$  can directly interact with ion channels. Muscarinic acetylcholine receptors of the type encoded by *CHRM1*<sup>294</sup> are responsible for the slowing of the heart upon stimulation by the vagus nerve. It is now understood that this effect is attributable to the binding of  $G\beta\gamma$  to the cytosolic side of the heteromeric channel *Kir3.1/3.4*.<sup>295</sup> This opens the channel (an inward rectifier) and produces the observed inhibition of the cardiac cell. In the brain this type of metabotropic inhibition is provided by homo- and heteromeric channels composed of *Kir3.1* and *Kir3.2* (encoded by *KCNJ3* and *KCNJ6* respectively). *Kir3.2* activation can also result from the metabotropic action of the serotonin receptor *5-HT<sub>1A</sub>* (encoded by *HTR1A*)<sup>272</sup>.

In other cases neither  $G\alpha\cdot\text{GTP}$  nor  $G\beta\gamma$  interacts with an ion channel directly but rather employs a *second messenger*. For example, adenylyl cyclases, such as those encoded by *ADCY2* and *ADCY8*, may be stimulated by either  $G\alpha\cdot\text{GTP}$  or  $G\beta\gamma$  to catalyze the formation of the second messenger cAMP. cAMP can have fairly direct effects on certain neuronal parts such as *HCN1*, the voltage dependence of which depends on cAMP concentration. For the most part, however, the effects of cAMP are mediated by cAMP-dependent protein kinase (PKA), a tetramer consisting of two regulatory and two catalytic subunits. The subunit encoded by *PRKAR2A* can fill one of the regulatory slots<sup>296</sup>. This pathway can be shut down by either phosphatases removing the phosphoryl groups from the target proteins or enzymes known as phosphodiesterases (PDEs), which hydrolyze cAMP to AMP (or cGMP to GMP). Seven PDEs are encoded by DEPICT-prioritized genes (*PDE1A*, *PDE1C*, *PDE2A*, *PDE4C*, *PDE7B*, *PDE10A*, *PDE11A*)<sup>297</sup>.

The phosphorylated membrane phospholipid  $\text{PIP}_2$  also plays a role in second-messenger systems. The one class of phospholipases that can be activated by G proteins is PLCB. *PLCBI* and *PLCB3* both encode PLCB enzymes. The reaction catalyzed by

PLCB cleaves PIP<sub>2</sub> into two products, IP<sub>3</sub> and DAG. (This is the biological significance of the cluster named after **DAG and IP<sub>3</sub> signaling**, which includes the gene sets **G-protein mediated events** and **PLC $\beta$  mediated events**.) The water-soluble IP<sub>3</sub> diffuses down into the cytoplasm and binds to receptors on the smooth endoplasmic reticulum. This binding results in the release of Ca<sup>2+</sup> into the cytoplasm, where it can influence ion-channel activity and a host of other functions.

Phosphatases are necessary to terminate the actions of kinases in timely fashion, and one of the specific phosphatases whose neuronal activities have perhaps been most well studied is PP2A (subunits of which are encoded by *PPP2R2A*, *PPP2R2D*, *PPP2R5B*, and *PPP2R5C*)<sup>298</sup>.

*Synaptic Communication and Plasticity: Multiple Spatiotemporal Scales of Synaptic Plasticity.* Glutamate receptors of both the AMPA and NMDA type are anchored to an electron-dense band of proteins lying immediately beneath the plasma membrane of the postsynaptic side. This band is known aptly enough as the *postsynaptic density* (PSD). The exemplary gene set **DLG4 PPI subnetwork**, the most statistically significant result in **Supplementary Table 8** ( $P = 4.64 \times 10^{-21}$ ), is defined by perhaps the most important protein in the PSD; the gene set **postsynaptic density** is a member of the corresponding cluster. Much remains to be learned about the formation and function of the highly complex PSD, but one of its roles is the provision of an interface between the channels riddling the plasma membrane and the cytoplasmic machinery that reshapes the signaling properties of the neuron. (It is now known that hundreds of distinct proteins contribute to the PSD. We will only highlight a handful encoded by genes that have been reasonably well studied in this context.)

If the PSD is regarded as a laminar structure, then the top layer consists of CAMs, neurotransmitter receptors, and TARPs<sup>299</sup>; the deepest layer is the interface with the cytoskeleton. The intermediate layer consists of scaffold proteins that lack enzymatic activity themselves but serve to fix the transmembrane proteins of the top layer in place and connect them to their more dynamic signaling targets in the dendritic spine. The fixing of glutamate receptors into clusters seems to be quite important: a cryptic columnar structure spanning the synaptic cleft, with a diameter of ~80 nanometers, encompasses both *RIMS1/2*-enriched active zones on the presynaptic side and receptor-enriched clusters on the postsynaptic side<sup>300</sup>. The main components of the intermediate PSD layer include the MAGUK, SHANK, and AKAP families of scaffold proteins.

A typical MAGUK has three PDZ domains, an SH3 domain, and inactive guanylate kinase domain<sup>301</sup>. Many synaptic transmembrane proteins (including glutamate receptors, voltage-gated ion channels, and CAMs) contain conserved PDZ-binding motifs in their

cytoplasmic tails that hook up with the PDZ domains of MAGUKs. DLG4 (also known as PSD95), probably together with **DLG2** (also known as **PSD93**), interacts in this way with the GRIN2 subunits of the NMDA-type glutamate receptor. **DLG1** (also known as **SAP97**) interacts more indirectly with the AMPA-type glutamate receptor, using TARPs related to **CACNG3** as intermediaries. **MAGI2** (also known as **SSCAM**) interacts with TARPs as well<sup>302</sup> and appears to regulate trafficking of those AMPA-type receptors that include a **GRIA2** subunit<sup>303</sup>.

**SHANK2** and **SHANK3** encode master scaffolding proteins with domains that interact with many other dendritic proteins, including NMDA-type glutamate receptors (via DLG4) and the **GRIA1** subunit of the AMPA-type receptor<sup>304</sup>. **SHANK3** is one of the genes most frequently found to be mutated in ASD patients<sup>305–307</sup>. A recent study found that normal sociability in **SHANK3**-deficient mice can be restored by inhibiting the kinase encoded by **CLK2**<sup>308</sup>.

AKAPs were originally defined by the anchoring of the kinase PKA to specific cellular locations. They are now recognized for their ability to form complexes integrating several distinct pathways. In particular, **AKAP5** (formerly called **AKAP79**) encodes another master scaffolding protein that links actin, MAGUKs, cadherins, kinases, phosphatases, glutamate receptors, and voltage-gated ion channels<sup>309</sup>. **AKAP5** anchors both the kinase PKA and the phosphatase calcineurin, potentially balancing these antagonists against each other. The protein encoded by **AKAP9** (formerly **AKAP450**) plays a similar co-localizing role, except that its phosphatase is PP1 rather than calcineurin.

*Long-term potentiation* (LTP) is mediated by the kinase CaMKII, a 12-mer with subunits encoded by **CAMK2A**, **CAMK2B**, and **CAMK2G**. This kinase is so abundant in neurons that it accounts for roughly one percent of all protein, suggesting a structural role in the PSD as well as an enzymatic one, and indeed **CAMK2A** can bind to the **GRIN2B** subunit of the NMDA-type glutamate receptor. Incoming Ca<sup>2+</sup> ions admitted through activated NMDA-type receptors bind to a gene product called calmodulin. The Ca<sup>2+</sup>/calmodulin complex, in turn, binds to the individual subunits of CaMKII and renders the complex into an active kinase.

The first step in the accumulation of more AMPA-type glutamate receptors seems to be the enlargement of the dendritic spine itself. CaMKII activation is the likely trigger of the pathways responsible for the increase and rearrangement of the cytoskeletal actin in a larger spine, although not all of the precise mechanisms have yet been elucidated. **DBN1** encodes an actin-binding protein that renders its substrate resistant to depolymerization, which may help stabilize the cytoskeleton in a newly enlarged spine. **ACTR2** encodes a

component of the **ARP2/3** complex, which is involved in the initiation of actin filament bifurcation<sup>310</sup>; the extension of the cytoskeleton into the spine has a branching structure. One study has found that knockdown of **ARP2/3** in mice leads to a failure of sustained spine enlargement and impairment of object memory<sup>311</sup>. One target of CaMKII is **KALRN**<sup>312</sup>, a catalyst of GDP/GTP exchange on GTPases acting on the actin cytoskeleton (e.g., **RHOA**). Alternative splicing of **KALRN** produces several isoforms, each of which may play a distinctive role in dendritic development and rearrangement<sup>313</sup>. The target of at least one isoform, **RHOA**, can spread into the parent dendritic shaft and adjacent spines within ~5 micrometers<sup>314</sup>. Fresh instances of these proteins and others must be supplied by the machinery of protein synthesis in the dendritic shaft, if spine enlargement is to be prolonged, and the signal to initiate this resupply can come from activation of the receptor tyrosine kinase **NTRK2** by **BDNF** of likely autocrine origin; the release appears to be triggered by a combination of glutamate and backpropagating action potentials<sup>315</sup>. This late-phase LTP can also spread to other stimulated spines on the same dendritic branch, up till tens of micrometers away<sup>316</sup>.

The insertion of new receptors appears to require substantial remodeling of the PSD. CaMKII mediates the disassembly of PSD proteins such as **DLG4** and **SHANK3**; other mechanisms presumably reinsert these scaffold proteins in the new configuration. **NLGN1** seems to be trafficked to the top layer of the PSD, where it they may be able to interact with **DLG4** to reduce the diffusion of new AMPA-type receptors in the plasma membrane and trap them in place. It is possible that the free AMPA-type receptors are numerous enough to constitute a pool that can be drawn upon for synapse enhancement without waiting for new arrivals to arrive via exocytosis. Knockdown of both **SYT1** and **SYT7** has recently been shown to abolish LTP, apparently by undermining a crucial step between  $\text{Ca}^{2+}$  influx and exocytosis of AMPA-type receptors<sup>317</sup>.

The processes of *long-term depression* (LTD) act to weaken a synapse<sup>318,319</sup>. One pathway leading to LTD is initiated through activation of the phosphatase calcineurin (the catalytic subunit of which is encoded by **PPP3CA**) by the  $\text{Ca}^{2+}$ /calmodulin complex. Calcineurin, in turn, dephosphorylates the product of **PPP1R1B** (also called **DARPP32**)<sup>320</sup>, which can then no longer inhibit the phosphatase PP1 (subunits of which are encoded by **PPP1CA**, **PPP1CB**, **PPP1CC**, **PPP1R9A**, **PPP1R21**, and **PPP1R35**)<sup>321</sup>. Freed from tonic inhibition, PP1 then dephosphorylates AMPA-type glutamate receptors, which subsequently admit less current when opened. Through mechanisms that are not well understood, dephosphorylation also leads to the removal of AMPA-type receptors from the dendritic spine via clathrin-mediated endocytosis and shrinkage of the spine head itself.

So far we have been discussing forms of LTP and LTD mediated by NMDA-type glutamate receptors, as befits the exemplary gene set **post NMDA receptor activation events**, but other forms are also operative. One form of LTD is mediated by metabotropic glutamate receptors, the receptor type depending on the brain region. For example, **GRM1** and **GRM2** mediate LTD at Purkinje synapses in the cerebellum and mossy-fiber synapses in the hippocampus respectively. The PSD actually connects **GRM5** to NMDA-type receptors, suggesting some kind of synergistic effect on synaptic plasticity of these two glutamate receptor types<sup>322–324</sup>. Both members of the  $\delta$  subfamily of ionotropic glutamate receptors are encoded by DEPICT-prioritized genes (**GRID1**, **GRID2**). Until recently their endogenous ligands were unknown and their functions not well understood, but now it has been shown that a transsynaptic complex beginning with a neurexin on the presynaptic side and ending with **GRID2** on the postsynaptic side can induce LTD at cerebellar synapses via endocytosis of AMPA-type glutamate receptors<sup>325</sup>. There is also the possibility that wholesale structural changes—e.g., the extension of new spines and the withdrawal of old ones, the corresponding swelling and subsidence of axon boutons—play an important role<sup>265</sup>. It is not known, however, whether such changes occur in the *in vivo* adult brain. There is much evidence, in fact, that dendritic arbors and spines are quite stable once the individual has reached early adulthood<sup>326</sup>.

*Comparison with Genes and Gene Sets Prioritized in a GWAS of Cognitive Performance.* To help put our findings in context, we compare our results to those of the largest published GWAS of *CP*<sup>327</sup> (that does not rely on a joint analysis with *EduYears*). We note that our study of *EduYears* identifies far more genes and gene sets than this GWAS of *CP*, and for that reason, we focus on our discussion on the extent to which the *CP* results replicate in our study. However, the GWAS of *CP* has extensive sample overlap with the current study of *EduYears*. Therefore the agreement between the two studies detailed below should not be interpreted as fully independent replication.

Both the *CP* study and the current *EduYears* study employed MAGMA to prioritize likely causal genes. Of the 47 genes found by the *CP* study to be Bonferroni significant after correction for 18,338 genes ( $P < 2.7 \times 10^{-6}$ ), 41 attained  $FDR < 0.05$  in our current GWAS of *EduYears* (**Supplementary Table 29**). The *CP* study considered any gene reaching the Bonferroni threshold times 10 ( $P < 2.7 \times 10^{-5}$  and  $P \geq 2.7 \times 10^{-6}$ ) to be suggestively significant. Of the 58 genes found by the *CP* study to be suggestively significant, 53 attained  $FDR < 0.05$  in our study of *EduYears*.

The *CP* study employed a use of MAGMA that we did not conduct in our own GWAS of *EduYears*—the testing of gene-set enrichment. This MAGMA procedure falls within a class of methods where statistical power increases very slowly with sample size<sup>328</sup>. We will nevertheless compare the results of this study with ours, which were



obtained by applying different tools. In the *CP* study, a single gene set was found to be significantly enriched, **regulation of cell development**<sup>327</sup>. This large and heterogeneous GO gene set overlaps with many others, and the study of *CP* reported that its significance was apparently driven by **regulation of nervous system development**, **negative regulation of dendrite development**, **neuron spine**, and **myelin sheath**<sup>bb</sup>. In our DEPICT results, **regulation of cell development** is highly significant ( $P = 1.18 \times 10^{-12}$ ), and furthermore it is a member of the cluster named after **regulation of nervous system development** (Supplementary Table 8). Other members of this cluster include **positive regulation of neuroblast proliferation**, **abnormal neuronal precursor proliferation**, **regulation of neurogenesis**, **regulation of neuron differentiation**, and **abnormal hippocampal commissure morphology**, indicating that the relevant biology involves the progression of the neuron over time—from its infancy as a progenitor, to its maturation into an information-processing network. Our application of the Affinity Propagation algorithm placed the gene set **dendritic spine** in the cluster named after **DLG4 PPI subnetwork**, and it is among the most significantly enriched gene sets in our study ( $P = 2.35 \times 10^{-17}$ ). As mentioned previously in our discussion of glial cells, **myelin sheath** is not significant in our results, but this was the gene set mentioned in the main text of Sniekers et al.<sup>327</sup> that overlaps least (14 percent) with **regulation of cell development**.

Many of these gene sets are also significant in our robustness analyses. **Regulation of cell development** was returned by PANTHER (2.19-fold enrichment, Bonferroni  $P = 8.78 \times 10^{-15}$ ; Supplementary Table 30), as were **regulation of nervous system development** (2.36-fold enrichment, Bonferroni  $P = 1.9 \times 10^{-18}$ ) and **dendritic spine** (3.19-fold enrichment, Bonferroni  $P = 3.99 \times 10^{-6}$ ). **Regulation of nervous system development** is one of binary gene sets showing the most enrichment according to stratified LD Score regression (1.56-fold enrichment,  $P = 2.54 \times 10^{-6}$ ; Supplementary Table 36), and **regulation of cell development** is not far behind (1.50-fold enrichment,  $P = 6.16 \times 10^{-6}$ ).

In short, the one significantly enriched gene set reported in the largest GWAS of *CP* to date is strongly implicated in our current study of *EduYears*. Related gene sets also attaining significance in our study indicate that one of the biological mechanisms driving the high genetic correlation between *CP* and *EduYears* is the cellular development of neural progenitors and postmitotic neurons.

---

<sup>bb</sup> In the current DEPICT inventory, **negative regulation of dendrite development** is absent. **Neuron spine** is a duplicate of **dendritic spine**; we retained the latter.

## 5.7. Results: Causal SNPs

*Stratified LD Score regression.* The results of estimating the stratified LD Score regression model are given in **Supplementary Table 31**. The top results by heritability enrichment refer to evolutionary conservation (**conserved [GERP RS > 4]**, 14.2-fold enrichment,  $P = 8.55 \times 10^{-8}$ ; **conserved [Lindblah-Toh]**, 9.22-fold,  $P = 3.16 \times 10^{-13}$ ). This is a typical result of stratified LD Score regression regardless of the phenotype<sup>1,18,24</sup>. A number of the significant annotations refer to histone marks, and the effects of these on gene regulation have already been discussed in the context of the gene-set enrichment analysis. As a whole, the baseline results show that annotations indicative of gene regulation (**transcription start site, weak enhancer**) tend to increase a SNP's effect on *EduYears*.

The results of estimating the effects of the fgwas annotations, one at a time in conjunction with the baseline model, are given in **Supplementary Table 32**. The nine annotations showing the most enrichment all refer to replicates of experiments assaying DNase I hypersensitivity in the fetal brain (**Supplementary Figure 23**). This finding replicates the finding from Okbay et al's smaller-sample GWAS of *EduYears*<sup>1</sup>, but our estimate of the magnitude of the enrichment is larger. The **fetal brain** annotation with the greatest effect size by this measure exhibits 12.5-fold enrichment ( $P = 1.63 \times 10^{-23}$ ); that is, whereas only about 2.5 percent of all 1000 Genomes common SNPs bear this annotation, they account for more than 30 percent of the heritable variance in *EduYears*. Moreover, all 12 instances of **fetal brain** exhibit at least 7.4-fold enrichment. When the measure of effect size is the annotation's partial regression coefficient, the top nine annotations remain instances of **fetal brain**. (The fgwas annotations show a correlation of 0.94 between the enrichment and coefficient measures of effect size.)

Other tissues/cell types in which the residence of a SNP in a DNase I hypersensitive region predicts a greater effect on *EduYears* include neural progenitor cells (7.0-fold,  $P = 3.86 \times 10^{-8}$ ) and the fetal spinal cord (8.3-fold,  $P = 1.43 \times 10^{-18}$ ). The top annotations referring to genome segmentations are defined mostly by active transcription start sites, as in embryonic stem cells (8.7-fold,  $P = 2.47 \times 10^{-4}$ ).

*CAVIARBF.* The optimal parameters  $\{\alpha, \lambda\}$  were  $\{0, 64\}$  (selected as described in Section 5.4), reflecting ridge regression with a cost parameter of  $\lambda = 64$ . We recovered 127 SNPs with PIP > 0.9 in 120 unique loci. Of these SNPs, 35 have annotations related to the fetal brain (22 with **DS16302**), 26 are conserved according to the **Lindblah-Toh** annotation, and 14 bear annotations of both these types. This result is consistent with our results from stratified LD Score regression, discussed above. Twenty-two of the 127



SNPs with PIP > 0.9 are coding SNPs, and eight of these are nonsynonymous. The regularized effect sizes of the annotations are available in **Supplementary Table 37**.

Our results remain robust to replacing our 1000 Genomes reference panel with the UKB reference sample. Adjusting for the smaller number of SNPs in this reference panel (308K vs. ~331K SNPs available for CAVIARBF), the number of prioritized SNPs we recover is very similar: 115 SNPs with PIP > 0.9 in 111 unique loci. In addition, the eight nonsynonymous SNPs that are prioritized using 1000 Genomes are also prioritized using the UKB reference sample. The robustness of our CAVIARBF fine mapping analysis to differences in reference sample may be due in part to the availability of functional annotation information, which reduces the importance of LD estimation precision by providing additional information that the model can use to differentially prioritize SNPs within each locus.

Several candidates for causal SNPs in **Supplementary Table 10** are particularly noteworthy. The most intriguing candidate is the nonsynonymous SNP rs61734410, which reaches genome-wide significance in our MTAG analyses of all four traits. It also reaches genome-wide significance in our GWAS analyses of *Highest Math* ( $P = 6.27 \times 10^{-14}$ ) and *Math Ability* ( $P = 2.97 \times 10^{-8}$ ). The SNP fails to reach genome-wide significance in the GWAS analysis of *CP* ( $P = 0.01$ ), but the sign of the effect remains concordant. The SNP resides in **CACNA1H** (**Supplementary Figure 9C**). As mentioned earlier, this gene encodes the pore-forming subunit of **Cav3.2**, which is found predominantly on dendrites. Cav3.1 and **Cav3.3** channels admit low-threshold and rapidly inactivating  $\text{Ca}^{2+}$  currents that promote rhythmical bursting of the neuron, but a recent study reported that **Cav3.2** differs from the other Cav3 channels in that it plays a role in the synaptic trafficking of NMDA-type glutamate receptors; away from synapses it appears not to be as strongly expressed in the mature rat brain as its family members<sup>329</sup>. *De novo* nonsynonymous mutations of this gene have been implicated as causes of childhood absence epilepsy through the mechanism of a lowered activation threshold<sup>330</sup>. The amino-acid substitutions induced by most of these mutations are located in a cytoplasmic loop between transmembrane domains I and II, bearing out the importance of this particular domain to neurophysiological function, and it happens that the substitution induced by rs61734410 also occurs in this domain.

**CACNA1H** is more highly expressed in the human brain prenatally<sup>51</sup> and is a high-ranking member of the exemplary gene sets **protein binding transcription factor activity**, **partial postnatal lethality**, **axon guidance (Reactome)**, and **regulation of neuron projection development**. Earlier we cited a recent study indicating that voltage-gated channels of the Cav3 type mediate GABA-induced formation of dendritic spines<sup>226</sup>, a function that is perhaps consistent with these genes sets. It is thus possible that the

nonsynonymous SNP rs61734410 affects *EduYears* through perturbations of both brain development and online neurophysiology.

One interesting trend is that several of our genes encoding transcription factors critical for neuron differentiation and prioritized in our earlier studies (*POU3F2*, *BCL11A*, *TBRI*)<sup>1,2,331</sup> are now linked to candidates for causal SNPs (rs62422687, rs10189857, rs11678980). These SNPs may also be particularly fruitful foci of follow-up research, because of their relatively large effect sizes and likely effects on *EduYears* through multiple mechanisms mediated by different target genes.

## 5.8. Omnigenicity

At the request of a referee, we have applied an analysis reported in a recent paper<sup>332</sup> to the current GWAS of *EduYears*. The authors of this work annotated each protein-coding gene in the human genome as **broadly expressed** or **specifically expressed** in one of several tissues. See pp. 7-9 of the Supplemental Information of their paper for the details of how these annotations were derived from the GTEx data<sup>333</sup>. For purposes of stratified LD Score regression, any SNP in an exon (or within 1 kb of either the first exon's start or the last exon's stop) inherited the tissue-level annotation of its encompassing gene. The paper analyzed schizophrenia. The paper found that despite showing less heritability enrichment than brain-expressed genes, broadly expressed genes account for more total heritability by virtue of being more numerous.

We obtained the assignment of genes to tissues from these authors and applied their procedure to the current *EduYears* summary statistics exactly as described. The results, shown in **Supplementary Figure 24** and **Supplementary Table 34**, are similar to those displayed in Figure 2B of Boyle et al.<sup>332</sup>. In the right panel of **Supplementary Figure 24**, we plot fraction of heritability rather than absolute amount of heritability because the former is the actual output of stratified LD Score regression. The absolute amount of heritability can be computed by multiplying the proportion by the total heritability reported in the log file, but doing so simply rescales the *x*-axis of **Supplementary Figure 24**.

When the tissues are ranked by heritability enrichment, the top result is **frontal cortex** (10.4-fold enrichment,  $P = 2.84 \times 10^{-5}$ ), followed by **pituitary** (7.65-fold enrichment,  $P = 0.08$ ) and **broadly expressed** (4.42-fold enrichment,  $P = 2.57 \times 10^{-4}$ ). When the tissues are ranked by proportion of heritability, however, **broadly expressed** (proportion of  $h^2 = 0.026$ , one-sided  $P = 6.25 \times 10^{-7}$ ) overtakes **frontal cortex** (proportion of  $h^2 = 0.019$ , one-sided  $P = 1.36 \times 10^{-6}$ ). If we assume (unrealistically) no sampling covariance between these two estimates, then the difference between **broadly expressed**

and **frontal cortex** in proportion of heritability is not significant ( $P = 0.30$ ). The remaining results combined do not account for even half as much heritability as **frontal cortex**.

Since some of the key concepts used by Boyle et al.<sup>332</sup> are not formally defined, there is some ambiguity as to what data would be needed to falsify the “omnigenicity” hypothesis. A possible interpretation of our results in line with the thesis of Boyle et al. might go as follows. Causal sites acting through **frontal cortex** genes may have direct and biologically interpretable effects on *EduYears*. The effects of **frontal cortex** genes, however, are likely to be moderated by **broadly expressed** genes acting through complex networks “with no direct relevance to [*EduYears*]” (p. 1184). And since **broadly expressed** seems to account for more heritability than **frontal cortex**, it may be that GWAS of high-frequency variants is no longer identifying “core genes” with “interpretable mechanistic links to [*EduYears*].” Such an impasse would motivate a turn to alternative designs, such as sequencing studies of low-frequency variants, to identify more core genes expressed specifically in the brain. Furthermore, more research into neuron-specific gene regulation is needed to understand how broadly expressed genes that are only indirectly relevant can nevertheless be associated with so much heritability.

We agree with much of this interpretation. It is reasonable to posit that there will be a point of diminishing returns, past which GWAS of a given trait will identify very few additional core biological processes and cell compartments. It is not clear whether we have yet reached this point with the current study of *EduYears*. Our results provide pointers to a vast sweep of neurobiology, ranging from early development to online neurophysiological function. We still lack clear evidence for the relevance of some neurobiology (segmentation and head induction during the first month after conception, glial function, transport of cargo through axons and dendrites), and increasing the sample size may bring these missing pieces into the picture very slowly. It is conceivable that sequencing studies of low-frequency variants, using aggregate tests to alleviate the lack of statistical power, could identify some of the missing neurobiology more efficiently. Finally, we agree that an improved understanding of gene regulation will enhance our ability to interpret existing GWAS results.

At the same time, we do not believe that it is necessarily appropriate to label genes that fail to qualify for the **frontal cortex** annotation as “indirectly relevant” or “uninterpretable,” at least not if the purpose of the labels is to convey to readers whether or not a given gene is a promising candidate for functional follow-up studies. For example, *CACNA1H*, *BCL11A*, and *POU3F2* are genes mentioned in our discussion of likely causal SNPs. None bear the **frontal cortex** (or the **broadly expressed**) annotation. Nevertheless, it is clear from the literature that when grossly perturbed, these genes have

substantial impacts on brain development and functioning (Supplementary Note section 5.6). Thus, while the **frontal cortex** annotation is certainly a reasonable factor to consider when prioritizing candidates for functional follow-up studies, other considerations are also reasonable to consider, such as the confidence with which the causal site responsible for the GWAS signal can be isolated, the effect size of the putative causal site, and what is currently known about the role of the gene in the brain (**Supplementary Table 10**). We suspect that, often, such case-specific knowledge or its absence may be more compelling than the general argument that expression in additional tissues renders a brain-specific mechanism more elusive.

## 6. Prediction

### 6.1 Introduction

In this section, we assess empirically how well various polygenic scores (PGS) derived from the analyses described in **Supplementary Section 1** predict a host of phenotypes related to educational attainment, academic achievement, and cognition in European-ancestry holdout samples. Additionally, we test the predictive power of our main educational attainment polygenic score in a sample of African-American individuals. We begin by describing and motivating the methodology used to generate all polygenic scores analyzed in this section. Next, we analyze polygenic scores based on summary statistics from our primary GWAS of *EduYears*. In the final subsection, we evaluate polygenic scores based on summary statistics from our remaining association analyses, including the joint analysis of *EduYears*, *CP*, *Math Ability*, and *Highest Math*.

### 6.2 Constructing Polygenic Scores

In general, a polygenic score for an individual is defined as a weighted sum of a person’s genotypes at  $J$  SNPs,

$$\hat{g}_i = \sum_{j=1}^J x_{ij} w_j. \quad (6.1)$$

Methodologies for PGS construction differ primarily across two dimensions: how to generate the weights  $w_j$ , and how to determine which  $J$  SNPs to include<sup>334,335</sup>. Since a single study rarely contains enough data to conduct a well-powered GWAS, researchers generally meta-analyze GWAS summary statistics across cohorts, and the meta-analyzed cohorts serve as the discovery sample. Standard methodologies for PGS construction only require access to GWAS summary statistics from a discovery sample, which are then applied to the individual genotype data in a smaller, holdout (prediction) sample.

A common and simple strategy for constructing polygenic scores from summary statistics is “pruning and thresholding” (P&T). With the P&T method, the weights are set equal to the coefficient estimates from univariate regressions of the phenotype on each variant  $j$ . The  $J$  SNPs are selected using a pruning algorithm that ensures the markers included in the score are all approximately independent of each other. The purpose of pruning is to eliminate genetic variants that are correlated with (i.e., are in linkage disequilibrium, or LD, with) already included variants. Including highly correlated variants would cause the PGS to “double count” the effect of a causal variant. To avoid overfitting, weights are estimated in a training sample that is independent of the prediction cohort. Sometimes, the set of SNPs is further restricted by omitting SNPs

whose  $P$  value for association with the phenotype is above a certain threshold. The purpose of this restriction is to try to boost the signal-to-noise ratio by only including genetic variants that are most likely to be truly associated with the phenotype in the PGS.

The analyses in this section are all based on scores derived using a Bayesian approach, LDpred estimation, that explicitly models and accounts for genetic architecture. LDpred has generally been found to perform better than the cruder P&T approach. We conduct a version of LDpred estimation that assumes a Gaussian prior for the distribution of effect sizes and sets the weight for each variant equal to the mean of its posterior distribution after accounting for LD. The theory underlying LDpred is derived assuming the variance-covariance matrix of the genotype data in the training sample is known. In practice, this matrix is not known, so we follow Vilhjalmsón et al.<sup>335</sup> and replace the training-sample variance-covariance matrix by an approximation that is estimated using observed LD patterns in a reference sample of conventionally unrelated individuals with European ancestry. Because long-range LD is assumed to be absent, the approximating matrix is block diagonal. In each of our prediction cohorts, we use cohort-specific genotype data as the reference sample (after dropping cryptically related individuals and ancestry outliers).

All prediction analyses reported below are based on summary statistics from association analyses conducted in independent discovery samples. We omit from the meta-analysis the three cohorts we use for prediction—*Add Health*, the *HRS*, and the *WLS*—which we discuss in more detail below. To make fairer comparisons across our prediction cohorts, we impute the genotypic data for all three prediction cohorts and then use only HapMap3 SNPs available in these imputed datasets to construct our scores. We use HapMap3 SNPs because these SNPs are generally well imputed and provide good coverage. Samples are composed of European-ancestry individuals whose genotype data meet standard quality-control thresholds. All scores are based exclusively on HapMap3 SNPs that meet the following conditions: (i) the variant has a call rate greater than 98% in the prediction cohort; (ii) the variant has a minor allele frequency (MAF) greater than 1% in the prediction cohort; and (iii) the allele frequency discrepancy between the meta-analysis and the prediction cohort does not exceed 0.15.

We calculate the posterior means using the software LDpred<sup>335</sup>. The final scores are then obtained using PLINK<sup>336</sup>, multiplying the genotype probabilities at each SNP by the corresponding estimated posterior mean and summing over all SNPs.

In what follows, we adopt the convention of defining each score by the source of the association statistics, GWAS or MTAG, followed by the name of the phenotype whose association statistics are being used. For example, *GWAS-EduYears* refers to a PGS

constructed using weights from the association statistics from the primary GWAS of *EduYears*, whereas MTAG-CP is a polygenic score whose weights are derived from the MTAG association statistics for CP.

### 6.3 Defining Prediction Accuracy

All prediction analyses are based on ordinary least squares regressions of a phenotype on the PGS and a set of controls. Unless otherwise noted, all regressions include the following basic set of controls: a full set of dummy variables for year of birth, an indicator variable for sex, a full set of interactions between sex and year of birth, and the first 10 principal components of the variance-covariance matrix of the genetic data. The principal components were estimated in a subset of conventionally unrelated individuals in each prediction cohort, using HapMap3 SNPs<sup>337</sup> with a minor allele frequency greater than 0.01.

To evaluate prediction accuracy, we use a simple, two-step process. First, we regress the phenotype on our set of controls without the PGS. Next, we rerun the same regression but with the PGS included. For quantitative phenotypes, our measure of predictive power is the incremental  $R^2$ : the difference in  $R^2$  going from the regression without the PGS to the regression with the PGS. For binary outcomes, we proceed similarly but calculate the incremental pseudo- $R^2$  from a Probit regression. To obtain 95% confidence intervals (CIs) around the incremental  $R^2$ 's, we perform a bootstrap with 1000 repetitions.

### 6.4 GWAS-*EduYears* Polygenic Score

The main analyses reported in this section were conducted among European-ancestry subjects in the National Longitudinal Study of Adolescent to Adult Health (*Add Health*)<sup>338</sup> and the Health and Retirement Study (*HRS*)<sup>339</sup>. (To avoid complicating the exposition, we postpone analysis of the Wisconsin Longitudinal Study (*WLS*) until Section 6.5, because the *WLS* has a truncated distribution of *EduYears*.) Both studies provide comparable measures of completed education. Whereas *Add Health* offers detailed measures of scholastic and cognitive achievement in adolescence, the *HRS* contains measures of several dimensions of cognitive functioning in older individuals. Following the suggestion of a referee, we also conducted a prediction analysis among African-American subjects in the *HRS*.

In the imputed *Add Health* data, there are a total of 1,217,312 HapMap3 SNPs. Of these, 1,211,662 SNPs have a call rate above 98% and 1,196,228 are available in the meta-analysis. 26,930 SNPs with minor allele frequency less than 1% were removed, leaving 1,169,298 SNPs that were used to construct the *Add Health* scores.

In the imputed *HRS* data, there are a total of 1,216,794 HapMap3 SNPs. Of these, 1,144,251 have a call rate above 98% and 1,127,758 are available in the meta-analysis. 23,077 SNPs with minor allele frequencies less than 1% were removed, leaving 1,104,681 SNPs that were used to construct the scores.

**Supplementary Table 38** provides a summary overview of the results from our prediction analyses as well as descriptive statistics for the phenotypes considered.

*EduYears*. We begin by examining how well the GWAS-*EduYears* score predicts our primary phenotype, *EduYears*, available in both *Add Health* and *HRS*. The results are shown in **Supplementary Figure 13**. We estimate that a one-standard-deviation increase in the PGS is associated with 0.79 extra years of schooling in *Add Health*. The associated incremental  $R^2$  is 12.7%. In the *HRS*, we estimate a coefficient of 0.84, with an incremental  $R^2$  of 10.6%.

To gauge the impact of expanding the sample size used in the discovery sample, it is useful to compare the results obtained here with results based on polygenic scores from previous GWAS of educational attainment. **Supplementary Figure 26** depicts how predictive power changes as the sample size for the *EduYears* meta-analysis has increased over time. To maximize comparability, all numbers reported in the figure are based on scores constructed using identical methods (namely, the methods described above) and a common set of SNPs (namely, all HapMap3 SNPs present in each of the four meta-analyses). In the figure, the  $x$ -axis measures the size of the discovery sample for the meta-analysis used to construct the scores, and the  $y$ -axis measures incremental  $R^2$  associated with the score. Polygenic scores derived from the first large-scale GWAS of educational attainment ( $N = 126,559$ )<sup>2</sup> explain 4.0% of the variation in *EduYears* for *Add Health* and 2.8% of the variation in *EduYears* for the *HRS*. The predictive power increases to 6.9% for *Add Health* and 5.7% for the *HRS* when summary statistics from Okbay et al.'s<sup>1</sup> discovery analyses are used ( $N = 293,723$ ) and further increases to 8.5% for *Add Health* and 6.7% for the *HRS* when the weights are based on Okbay et al.'s<sup>1</sup> combined meta-analysis of discovery and replication cohorts ( $N = 405,072$ ). Our current meta-analysis with  $N = 1,131,881$  individuals increases prediction accuracy in both samples, to 12.7% in *Add Health* and 10.6% in the *HRS*. Taken as a whole, **Supplementary Figure 26** confirms the strong positive relationship between the sample size of a meta-analysis and the predictive power of a polygenic score created from a meta-analysis.

In **Supplementary Figure 10**, we measure the impact of increasing the  $P$  value inclusion threshold for the number of SNPs used to construct our *EduYears* polygenic score. The  $x$ -axis measures an increasing  $P$  value threshold for scores made with the



“pruning and thresholding” (P&T) method described above. These thresholds are  $5 \times 10^{-8}$ ,  $5 \times 10^{-5}$ ,  $5 \times 10^{-3}$ , and 1 (i.e., all SNPs). We also include, as a point of comparison, the LDpred score for our meta-analysis, constructed with HapMap3 SNPs and a fraction of causal SNPs set equal to 1 (i.e., the LDpred software is run using the spike-and-slab option with the fraction of causal SNPs equal to 1). The y-axis measures the incremental  $R^2$  associated with each score. Polygenic scores made using only the genome-wide significant SNPs ( $P$  value threshold  $5 \times 10^{-8}$ ) explain 3.8% of the variance in *EduYears* for *Add Health* and 2.5% for *HRS*. As the  $P$  value threshold increases to include more SNPs in the polygenic score, the predictive power also increases. Including all SNPs results in scores that explain 8.7% of the variance in *EduYears* for *Add Health* and 7.9% of the variance for *HRS*. Finally, the LDpred scores for *EduYears* generate further gains in predictive power for both *Add Health* and the *HRS*, to 12.7% and 10.6%, respectively.

*Other Education Outcomes in Add Health and the HRS.* In additional analyses of both the *Add Health* and *HRS* samples, we consider three binary variables related to educational attainment: (i) High School Completion, (ii) College Completion, and (iii) Grade Retention (i.e., retaking a grade). In both prediction cohorts, the PGS is significantly associated with all three outcomes. As shown in **Supplementary Table 38**, in *Add Health*, a one-standard-deviation increase in the score is associated with a 4.7 percentage-point increase in the probability of completing high school (incremental pseudo- $R^2 = 6.2\%$ ), a 15.6 percentage-point increase in the probability of completing college (incremental pseudo- $R^2 = 9.5\%$ ), and a 7.1 percentage-point reduction in the probability of having retaken a grade (incremental pseudo- $R^2 = 4.0\%$ ). These effects are substantial relative to the baseline prevalences of 93%, 33% and 18%, respectively. In the *HRS*, a one-standard-deviation increase in the score is associated with approximately an 8.8 percentage-point increase in the probability of completing high school (incremental pseudo- $R^2 = 6.0\%$ ), a 12.6 percentage-point increase in the probability of completing college (incremental pseudo- $R^2 = 8.3\%$ ), and a 4.3 percentage-point reduction in the probability of having retaken a grade (incremental pseudo- $R^2 = 1.9\%$ ). Again, these effects are substantial relative to the respective baseline prevalences of 82%, 25% and 14%.

In **Figure 4a** and **Supplementary Figure 11**, we show the mean frequencies of each of our 3 binary outcomes across quintiles of our *EduYears* polygenic score. First, we divide our polygenic score into five quintiles in both *Add Health* and *HRS* so that the 1<sup>st</sup> quintile reflects the lowest polygenic scores and the 5<sup>th</sup> quintile reflects the highest polygenic scores. Each quintile contains roughly 955 individuals in *Add Health* and roughly 1,720 individuals in the *HRS*. Then, for high school completion, college completion, and grade retention, we plot the mean prevalence in each quintile for both

*Add Health* and for the *HRS*. Using college completion as an example, depicted in **Figure 4a**, note that the mean prevalence increases substantially moving from the lowest quintile (11.6% in *Add Health* and 9.2% in the *HRS*) to the highest quintile (57.0% in *Add Health* and 44.7% in the *HRS*).<sup>cc</sup> Comparing the 1<sup>st</sup> and 5<sup>th</sup> quintiles, there is a 45.4-percentage-point difference in college completion in *Add Health* and a 35.5-percentage-point difference in the *HRS*.

*Cognitive and Academic Achievement Outcomes in Add Health.* In additional analyses in *Add Health*, we examine the relationship between the polygenic score and several phenotypes related to cognition and academic achievement. In the first wave of *Add Health*, when participants were 12–20 years old, verbal cognition was measured with a modified version of the Peabody Picture Vocabulary Test<sup>340</sup>. In this test, an interviewer reads a word aloud, and a respondent selects the illustration that best fits the word’s meaning. Eighty-seven items were included on this computer-adapted test, and scores were age-standardized.

In the third wave of *Add Health*, transcripts were collected from respondents’ high schools. From the transcripts, grade point averages (GPAs) were calculated using the common United States 0.0 to 4.0 range, both for Overall GPA and for subject-specific GPAs. We analyze Overall GPA, Math GPA, Science GPA, and Verbal GPA. In our GPA analyses, we control for high school fixed effects, so the estimates should be interpreted as measures of the extent to which within-school differences in PGSs are predictive of within-school differences in scholastic outcomes. For each of these cognitive and academic achievement phenotypes, we present the regression estimates, samples sizes, and incremental  $R^2$  values in **Supplementary Table 38** and **Supplementary Figure 13**.

The PGS is significantly predictive of all of our cognitive and academic achievement outcomes. In our analyses of the Peabody test scores, we find that the incremental  $R^2$  of the PGS is 7.5%. A one-standard-deviation increase in the score is associated with an increase in Overall GPA, Math GPA, Science GPA, and Verbal GPA of approximately 0.26, 0.27, 0.29, and 0.27, respectively. For Overall, Math, Science, and Verbal GPA, the incremental  $R^2$  is respectively 9.2%, 7.1%, 8.0%, and 7.4%.

*Cognitive Outcomes in the HRS.* In additional analyses of the *HRS* sample, we examined the relationship between the *EduYears* PGS and a number of cognitive phenotypes. Our first cognitive phenotype, *Total Cognition*, is the sum of four cognitive

---

<sup>cc</sup> The differences in high school and college completion rates across *Add Health* and the *HRS* are expected, given demographic trends in educational attainment over time in the US. We would expect respondents in the *HRS*, who were born in earlier years than respondents in *Add Health*, to have completed high school or college less often than *Add Health* respondents (and to have completed fewer years of education overall).

measures common across waves 3 through 10: an immediate word recall task, a delayed word recall task, a naming task, and a counting task, with a total score ranging from 0 to 35. Our second cognitive phenotype, *Verbal Cognition*, measures the subject's ability to define five words. Each definition supplied is rated as incorrect (0), partially correct (1) or completely correct (2), resulting in a total score ranging from 0 to 10. Third, to evaluate changes over time, we studied wave-to-wave changes in Total Cognition and Verbal Cognition,  $(x_t - x_{t-1})$ . Our fifth cognitive outcome, Alzheimer's, is an indicator variable equal to one for subjects who report having been diagnosed with Alzheimer's disease, and 0 otherwise. Our decision to include this variable was motivated by evidence that *EduYears* and Alzheimer's disease are modestly genetically correlated<sup>1</sup>.

Since the *HRS* data are longitudinal, the unit of analysis for our 4 cognitive outcomes is a person-year. For these analyses, because individual  $i$  took the cognitive tests at different ages, in our set of controls we replace our person-specific age variable with age at assessment (which differs for individual  $i$  across the cognitive outcomes). Furthermore, since the unobserved determinants of individual  $i$ 's outcome across waves are unlikely to be independent, we cluster all standard errors at the person level<sup>341</sup>.

For each of these cognitive phenotypes, we present the regression estimates, samples sizes, and incremental  $R^2$  values in **Supplementary Table 38**. The PGS is significantly predictive of the Total Cognition and Verbal Cognition scores (incremental  $R^2 = 2.7\%$  and  $4.7\%$ , respectively). The score is not significantly associated with wave-to-wave changes in *Total Cognition* or *Verbal Cognition*, nor is it significantly associated with our Alzheimer's measure.

*Chromosome Length.* In **Supplementary Figure 27**, we examine the relationship between chromosome length and the predictive power of an *EduYears* polygenic score. Each data point in the figure corresponds to a chromosome, with the  $x$ -axis coordinate denoting its length<sup>342</sup>, and the  $y$ -axis denoting the incremental  $R^2$  of a PGS based exclusively on genetic variants located on the chromosome. In both panels (*Add Health* on the left and *HRS* on the right), the dashed line is the best fit from a regression of the incremental  $R^2$  of each score on chromosome length, with the intercept constrained to zero. There is a strong, positive, linear relationship between the length of each chromosome and the predictive power of the by-chromosome polygenic score. The correlation between chromosome length and by-chromosome incremental  $R^2$  is 0.83 (SE = 0.125) in *Add Health* and 0.87 (SE = 0.109) in the *HRS*. For educational attainment, the more genetic data that a chromosome contains, the larger the predictive power of that chromosome's polygenic score, consistent with the idea that educational attainment is highly polygenic. **Supplementary Figure 28** shows by-chromosome incremental  $R^2$ s

from a sample-size weighted average of the *Add Health* and *HRS* results. The correlation between chromosome length and by-chromosome incremental  $R^2$  is 0.90 (SE = 0.098).

*Benchmarking the Predictive Power of the GWAS-EduYears PGS.* To benchmark the polygenic score's predictive power, we compared the incremental  $R^2$  of the polygenic score to that of several commonly used variables: mother's education, father's education, both mother's and father's education, verbal cognition, household income, and a binary indicator for marital status.

We conducted two separate analyses. In the first, we compared each variable individually to the PGS. Specifically, we calculated, for each variable, its incremental  $R^2$  using the same procedures as those described above for the PGS. The results, shown in Panel A of **Supplementary Table 39**, are broadly similar in *Add Health* and *HRS*. Panel A of **Supplementary Figure 12** shows the sample-size-weighted mean across the two cohorts.<sup>dd</sup> In both datasets, the PGS is a better predictor of educational attainment than marital status. The PGS is a marginally stronger predictor of *EduYears* than household income (a commonly used proxy for socioeconomic status) and a marginally weaker predictor than mother's education, father's education, or the respondent's score on a test of verbal skills. Finally, mother's and father's education together remain a substantially stronger predictor of own education than the PGS.

In many potential uses of PGSs as control variables, the relevant measure of predictive power is the incremental  $R^2$  of the PGS beyond variables that have already been measured. For example, consider an investigator performing a randomized evaluation of an intervention designed to improve educational outcomes. Suppose that prior to the intervention, the investigator obtained information about the demographic characteristics in **Supplementary Table 39**. Controlling for a PGS would reduce statistical uncertainty about the causal impact of the intervention, but the magnitude of the gain in precision depends on the incremental  $R^2$  of the PGS conditional on the set of covariates that are already available to the investigators (for a formal analysis, see Section 8 of the SOM of Rietveld et al.<sup>2</sup>).

In Panel B of **Supplementary Table 39**, we evaluate the incremental  $R^2$  of the polygenic score in predicting *EduYears* under various assumptions about the set of available covariates. The first row shows that the (sample-size-weighted mean) predictive power of our EA polygenic score is 11.4% when the only available covariates are sex, year of birth, and their interactions. Next, we control for a number of additional demographic variables one at a time: marital status, household income, mother's

---

<sup>dd</sup> In the last column for all analyses in **Supplementary Table 39**, we again use sample-size weighting to meta-analyze the results between *Add Health* and *HRS*.

education, and father's education. Controlling for marital status does not attenuate the incremental  $R^2$  of the polygenic score appreciably, whereas controlling for household income reduces its incremental predictive power to 9.3%. Controlling for mother's or father's education reduces the incremental  $R^2$  to 6.8% or 6.4%, respectively, and controlling for both mother's and father's education reduces the incremental  $R^2$  to 5.5%. Finally, when the full set of demographic characteristics is controlled for, the incremental  $R^2$  of the PGS is 4.6%. For a graphical illustration of these results, see Panel B of **Supplementary Figure 12**.

*EduYears* in African-American subjects in the HRS. We examined how well an *EduYears* polygenic score predicts *EduYears* among 1,519 African-ancestry individuals in the HRS. We constructed the score using the same LDpred weights as we used for European-ancestry subjects, described above. In order to identify the African-ancestry subjects, we used the list provided by the HRS, containing the family and individual identifiers of subjects identified as genetically of African-American ancestry. In the prediction analysis, we included the same set of age and sex controls as the European-ancestry analysis, and the first 10 principal components of the variance-covariance matrix of the genetic data of the African-ancestry sample. We found that the LDpred score predicts 1.6% (95% CI: 0.7% to 3.0%) of the variance in *EduYears* among African-ancestry individuals. This represents 85% attenuation in the predictive power of the score compared to the incremental  $R^2$  of 10.6% in our European-ancestry sample from HRS.

To provide context for this degree of attenuation, we identified three existing papers that examined the predictive power in an African-ancestry sample of a polygenic score constructed using weights from a European-ancestry GWAS. Note that the results from these papers are not entirely comparable to ours because they use different methods for estimating weights for the polygenic score and different methods of evaluating predictive power. One paper<sup>30</sup> examined a polygenic score estimated from the first large-scale GWAS of *EduYears*<sup>2</sup> ( $N = 126,559$ ). The prediction sample was Add Health. The measure of predictive power was the squared correlation ( $r^2$ ) between *EduYears* and the polygenic score, a measure that necessarily produces larger numbers than our measure of incremental  $R^2$ . They estimated  $r^2 = 0.032$  in their European-ancestry prediction sample ( $N = 8,630$ ) but only  $r^2 = 0.012$  in their prediction sample of African Americans ( $N = 3,456$ )—an attenuation of 63%.

A second paper<sup>343</sup> studied a binary phenotype, psychosis, using weights estimated in a GWAS with 150,064 individuals<sup>10</sup>. Unlike the other two papers that used U.S.-based prediction samples, this paper examined a U.K.-based prediction sample, the Biomedical Research Centre for Mental Health Genetics and Psychosis Study. The measure of predictive power was incremental Nagelkerke's  $R^2$ , where the control variables were 10

PCs and DNA origin (blood or buccal). In their European-ancestry prediction sample ( $N = 405$ ), the incremental Nagelkerke's  $R^2$  was 9.4%, while in their African-ancestry prediction sample ( $N = 3,456$ ), it was 1.1%. This represents an attenuation of 88%.

The third paper<sup>344</sup> studied BMI, and weights were obtained from a GWAS in a sample of 249,796 individuals<sup>345</sup>. The prediction sample was the Atherosclerosis Risk in Communities Cohort. The measure of predictive power was the incremental  $R^2$ , where the baseline set of controls was a measure of educational attainment. They estimated incremental  $R^2$  of 1.5% in their European-ancestry prediction sample ( $N = 8,286$ ) but only 0.22% in their prediction sample of African Americans ( $N = 2,442$ )—an attenuation of 85%.

We are cautious in comparing our results with these other results because of the differences in methodology and (in two out of three cases) phenotype, and we note that all of the attenuation estimates have uncertainty that we have not quantified. Nonetheless, our reading of this evidence is that the attenuation we observe in our African-American prediction sample is not out of line with what has been observed in other studies.

## 6.5 MTAG-Based Polygenic Scores

Next, we evaluated the predictive power of polygenic scores based on summary statistics from other association analyses we conducted, including the MTAG analysis of *EduYears*, *CP*, *Math Ability*, and *Highest Math*. Examining the predictive power of MTAG-based polygenic scores is of interest because Turley et al.<sup>13</sup> show theoretically that, very generally, MTAG-based scores are expected to outperform scores constructed from the underlying GWAS results. Empirically, previous studies have generally confirmed this expectation. For example, Turley et al.<sup>13</sup> apply MTAG to summary statistics from GWAS of depressive symptoms, neuroticism, and subjective well-being and find that the incremental  $R^2$  of their MTAG-based scores outperform the conventional scores by about 25%. A previous application of MTAG<sup>346</sup> to educational attainment, household income, and cognitive performance similarly reports that an MTAG-*CP* PGS has an incremental  $R^2$  of 6.9%, a substantial increase over the (sample-size weighted) mean of 2.5%<sup>327</sup> or 4%<sup>347</sup> reported in the largest-to-date GWAS of *CP*.

In this subsection, we compare the predictive power of GWAS-based and MTAG-based polygenic scores for both *EduYears* and *CP*. We then compare the gains in predictive accuracy from MTAG to the gains that would be expected theoretically when the MTAG assumptions are approximately accurate. Since MTAG is supposed to generate association results that are specific to each phenotype, we conclude by

evaluating in our application whether the “own-phenotype” MTAG-based polygenic scores in fact have greater predictive power than the “other-phenotype” MTAG-based polygenic scores.

We continue to use *Add Health* and the *HRS* for evaluating the predictive power of polygenic scores for *EduYears*. In our analyses of *CP*, we used *Add Health* and the Wisconsin Longitudinal Study (*WLS*) as prediction cohorts because, as discussed below, both have high-quality measures of *CP* that are likely to have a high genetic correlation with the phenotype used in our GWAS of *CP*.

*EduYears*. We began by rerunning some of the key analyses from the previous subsection, but with MTAG-based scores in lieu of GWAS-based scores. In both prediction cohorts, the incremental  $R^2$  of the MTAG-based scores is approximately one-half percentage point greater than the estimates from GWAS-based scores (**Supplementary Table 42**). In *Add Health*, the MTAG- and GWAS-based scores have incremental  $R^2$ 's of 13% and 12.7%, respectively. In *HRS*, the corresponding figures are 11.2% and 10.6%.

For completeness, we also repeat the same analysis in *WLS* and found incremental  $R^2$ 's of 6.8% (95% CI 6.2% to 7.2%) and 7.0% (95% CI 6.2% to 7.4%) for GWAS-*EduYears* and MTAG-*EduYears*, respectively. These estimates, however, must be interpreted in light of the fact that the original *WLS* study population, by design<sup>348</sup>, was drawn from the most educated part of the population, consisting almost exclusively of high school graduates (the intended sample was the set of all high school graduates from Wisconsin born between 1938 and 1940). In our *WLS* estimation sample, 98% of respondents completed high school, compared to the state-level average of 75% in the relevant birth cohort<sup>349</sup>. Adjusted for the range restriction, the *WLS* estimates are consistent with the *Add Health* and *HRS* results: if we impose the same truncation in *Add Health* or *HRS*—and limit the estimation samples to respondents who completed high school—the predictive accuracy of the GWAS-*EduYears* scores fall by approximately 20%. We also calculated the theoretically expected amount of attenuation under the assumption that the (PGS, *EduYears*) distribution in *WLS* is a truncated bivariate normal, with *EduYears* truncated from below at the 25<sup>th</sup> percentile. Under these assumptions, a correlation of 0.26 ( $\sqrt{0.068} \approx 0.26$ ) in the truncated distribution implies a correlation of 0.34 in the underlying (truncation-free) binormal distribution<sup>350</sup>. Hence, the restriction-of-range adjusted estimate for the predictive power of the GWAS-*EduYears* PGS is  $R^2 \approx 0.34^2 \approx 11.6\%$ , which is in the same ballpark as our *Add Health* and *HRS* estimates.

*Cognitive Performance*. Next, we examined how well MTAG-based and GWAS-based scores for *CP* and *EduYears* predict cognitive performance. We conducted these

analyses in two prediction cohorts: *Add Health* and *WLS*. (We omit the *HRS* even though it has a high-quality measure of cognitive functioning because the test was designed to measure risk for dementia and cognitive impairment in the elderly, and consequently a substantial fraction of the items measure learning and memory<sup>351</sup>.)

In the imputed *WLS* data, there are a total of 1,217,039 HapMap3 SNPs. Of these, 1,211,685 SNPs have a call rate above 98% and 1,200,740 are available in the meta-analysis. We drop the 29,981 SNPs with minor allele frequency less than 1%, leaving 1,170,759 SNPs that were used to construct the *WLS* scores.

In *Add Health*, our measure of cognitive performance is again the respondent's age-adjusted score on the Peabody Picture Vocabulary test. In the *WLS*, our measure of cognitive performance is the respondent's raw score on a Henmon-Nelson test of mental ability<sup>352</sup>. The Henmon-Nelson test is a 30-minute multiple-choice test that consists of 90 individual verbal or quantitative items. It is a psychometrically validated test whose scores are known to correlate highly with *g*. It is therefore reasonable to expect that the Henmon-Nelson test scores are highly genetically related to the COGENT and *UKB* cognitive measures on which the GWAS discovery-sample results for *CP* are based. Test scores are available for most *WLS* respondents.

In each of the two *CP* prediction cohorts, we compare the prediction accuracies of four polygenic scores: MTAG-*CP*, GWAS-*CP*, MTAG-*EduYears*, and GWAS-*EduYears*. The results are shown in Panel B of **Supplementary Table 42** and in **Figure 4C**. In *Add Health* the GWAS-*EA* and MTAG-*EA* scores outperform the GWAS-*CP* and MTAG-*CP* scores in predicting *CP*. However, the MTAG-*CP* score is more predictive than the GWAS-*CP* score, with an incremental  $R^2$  of 6.9% and a gain over the GWAS-*CP* score of 1.8%. In *WLS*, the MTAG-*CP* score is the most predictive of the four scores, with an incremental  $R^2$  of 9.7% and a gain over the GWAS-*CP* score of 2.7%.

## 6.6 Comparing Observed Gains in Prediction Accuracy to Theoretical Predictions

Next, we utilized a previously described framework<sup>13</sup>, to compare the observed prediction accuracies of the MTAG-based scores to their theoretically expected accuracies.

In a first step, we used a well-known theoretical result<sup>334</sup> to infer a SNP-based heritability for each trait:

$$R^2 = \frac{h^4}{h^2 + \frac{M_{\text{eff}}}{N}}, \quad (6.2)$$



where  $R^2$  is the predictive power of the GWAS-based PGS in the prediction sample,  $h^2$  is the SNP heritability of the phenotype in the prediction sample,  $N$  is the sample size used to create the PGS, and  $M_{\text{eff}}$  is the effective number of independent SNPs in the sample (assumed to be 60,000)<sup>334</sup>. This calculation gives an estimate of each trait's heritability specific to each prediction sample.

Next, we calculate the expected predictive power by replacing  $N$  with a GWAS-equivalent sample size calculated as follows:

$$N_{\text{equiv}} = N \frac{\overline{\chi_{\text{MTAG}}^2} - 1}{\overline{\chi_{\text{GWAS}}^2} - 1}, \quad (6.3)$$

where  $\overline{\chi_{\text{MTAG}}^2}$  and  $\overline{\chi_{\text{GWAS}}^2}$  are the average inflation-adjusted  $\chi^2$ -statistics. Finally, we calculate the difference between the expected predictive power of an MTAG-based score and a GWAS-based score and compare this difference to the one we observed empirically.

For *EduYears*, these calculations suggest that, relative to the GWAS-*EduYears* score, the MTAG-*EduYears* score will improve prediction accuracy of 0.4% in both *Add Health* the *HRS*. The observed gains are 0.3% and 0.6%. For *CP*, analogous calculations yield a theoretical prediction of 1.9% in both *Add Health* the *WLS*. The observed gains are 1.8% and 2.7%. In all cases, the observed gains are similar in magnitude to, and never statistically distinguishable from, those predicted by theory.

## 6.7 Comparing Trait-Specific Scores

In our final analysis, we compared the predictive power of the four MTAG-based scores for *EduYears*, *CP*, and *Math GPA*. Under the MTAG assumptions, an MTAG-based score constructed from own-phenotype association statistics should have greater predictive power for that phenotype than an MTAG-based score constructed from association statistics for one of the other phenotypes. For example, the MTAG-*EduYears* score should be a better predictor of *EduYears* than the MTAG-*CP*, MTAG-*Math Ability*, and MTAG-*Highest Math* scores. For *EduYears* and *CP*, it is straightforward to test empirically whether the observed rank ordering is consistent with theoretical expectations. For the two mathematics variables, we do not have comparable phenotypes in any of our prediction cohorts. We therefore used data from *Add Health* about the respondent's high-school GPA in math, which we consider to be the best available proxy. For the proxy, we expect the MTAG-*Math Ability* and MTAG-*Highest Math* scores to have relatively greater predictive power.

In **Supplementary Table 43** and in **Supplementary Figure 29**, we report the predictive power of each MTAG score for *EduYears* (both cohorts), *CP* (both cohorts), and Math GPA (*Add Health* only). We discuss each phenotype in turn, beginning with *EduYears*. In both cohorts, the MTAG-*EduYears* score is the best predictor of *EduYears*, followed by MTAG-*Highest Math*, MTAG-*CP*, and MTAG-*Math Ability* scores (though not shown, we observed an identical rank ordering in *HRS*). In *WLS*, the MTAG-*CP* score is the best predictor of Henmon-Nelson IQ, followed by the MTAG-*EduYears*, MTAG-*Highest Math*, and MTAG-*Math Ability* scores. In *Add Health*, the results are similar, albeit with the MTAG-*EduYears* and MTAG-*CP* scores having very similar predictive accuracies.

We also find that the MTAG-*EduYears* and MTAG-*Highest Math* scores are the best predictors of Math GPA, followed by MTAG-*CP* and MTAG-*Math Ability*. The MTAG-*Highest Math*, and MTAG-*Math Ability* scores are relatively better predictors of our proxy. The final columns of **Supplementary Table 43** report the results from additional analyses of *Verbal GPA* and *Science GPA*. For these variables, the predictive power of the MTAG-*Math Ability* and MTAG-*Highest Math* scores are lower, both in an absolute sense and relative to the MTAG-*EduYears* and MTAG-*CP* scores.

Overall, phenotype-specific MTAG scores are the best predictors of *CP* and *EduYears*. While we cannot make definitive statements about the two 23andMe math phenotypes, we find that MTAG scores based on them perform relatively better when the target phenotype is Math GPA, the best available proxy available to us.

## 7. Contributions and Acknowledgements

### **7.1 Author Contributions**

Daniel Benjamin, David Cesarini, Patrick Turley, and Peter Visscher designed and oversaw the study.

Aysu Okbay was the study's lead analyst, responsible for quality control and meta-analyses.

Analysts who assisted Aysu in major ways include: Edward Kong (quality control), Omeed Maghzian (COJO, MTAG, quality-control), Tuan Anh Nguyen-Viet (figure preparation), Hui Li (quality control), Chanwook Lee (quality control), Julia Sidorenko (UKB association analyses), and Richard Karlsson Linnér (UKB association analyses).

Peter Bowers and Edward Kong conducted the within-family association analyses.

The cross-cohort heritability and genetic-correlation analyses were conducted by Robbee Wedow and Meghan Zacher.

The analyses of the X-chromosome in UK Biobank were conducted by Julia Sidorenko; Aysu Okbay ran the meta-analysis.

James Lee organized and oversaw the bioinformatics analyses, with assistance from Tõnu Esko, Edward Kong, Kevin Thom, Tune Pers, and Pascal Timshel.

Polygenic-prediction analyses were designed and conducted by Aysu Okbay, Kevin Thom, and Robbee Wedow.

Besides the contributions explicitly listed above, Tushar Kundu, Ruoxi Li, and Rebecca Royer conducted additional analyses for several subsections. Chelsea Watson helped with coordinating among the participating cohorts.

Jonathan Beauchamp, Dalton Conley, Tõnu Esko, Magnus Johannesson, James Lee, Philipp Koellinger, David Laibson, Steven Lehrer, Sven Oskarsson, Matthew Robinson, Kevin Thom, and Jian Yang provided helpful advice and feedback on various aspects of the study design.

All authors contributed to and critically reviewed the manuscript. Edward Kong, James Lee and Robbee Wedow made especially major contributions to the writing and editing.

## 7.2 Cohort Contributions

Cohort	Author	Study Managemt	Data Collection	Genotyping	Genotype Prep.	Phenotype Prep.	Data Analysis	Writing
23andMe	Aaron Kleinman						X	
23andMe	Joyce Y. Tung	X						
23andMe	Nicholas A. Furlotte						X	
Add Health	Kathleen Mullan Harris	X	X					
Add Health	Jason D. Boardman		X					
COGENT	Joey W. Trampus	X	X		X	X	X	
COGENT	Todd Lencz	X	X				X	
COGENT	Anil K. Malhotra	X	X					
COGENT	Max Lam Zhan Yang				X	X	X	
EGCUT4	Tõnu Esko	X	X	X			X	X
EGCUT4	Andres Metspalu	X	X					
EGCUT4	Reedik Mägi				X		X	
EGCUT4	Lili Milani	X		X	X			
EGCUT4	Maris Alver					X	X	
Fenland	John R. B. Perry	X				X		
Fenland	Jian'an Luan	X			X	X		
Fenland	Claudia Langenberg				X		X	
Fenland	Nicholas J. Wareham	X			X	X		
Geisinger	Christopher F. Chabris	X				X	X	X
Geisinger	Shefali Setia Verma	X		X	X	X	X	X
Geisinger	Marylyn D. Ritchie	X	X	X	X		X	
GSII	Blair H. Smith		X					
GSII	David J. Porteous	X	X	X				
GSII	Caroline Hayward			X	X	X		
Norfolk	Ken K. Ong	X				X		
Norfolk	Nicholas J. Wareham	X			X	X		
Norfolk	Jing Hua Zhao				X		X	
Norfolk	Felix R. Day						X	
UKHLS	Meena Kumari	X						
UKHLS	Yanchun Bao						X	
UKHLS	Melissa C. Smart					X		
VIKING	James F. Wilson	X	X	X		X		
VIKING	David W. Clark						X	
VIKING	Peter K. Joshi				X	X	X	
VIKING	Harry Campbell		X	X				
WLS	Pamela Herd		X	X	X			
WLS	Jeremy Freese		X	X	X	X		

## 7.3 Additional Acknowledgements

**23andMe, Inc.** — 23andMe research participants provided informed consent to take part in this research under a protocol approved by the AAHRPP-accredited institutional review board, Ethical and Independent Review Services. We would like to thank the research participants and employees of 23andMe for making this work possible.

**Add Health (National Longitudinal Study of Adolescent to Adult Health)** — The National Longitudinal Study of Adolescent to Adult Health (*Add Health*) is supported by grant P01 HD031921 to Kathleen Mullan Harris from the Eunice Kennedy

Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. *Add Health* GWAS data were funded by NICHD grants to Harris (R01 HD073342) and to Harris, Boardman, and McQueen (R01 HD060726). For information about access to the data from this study, contact [addhealth@unc.edu](mailto:addhealth@unc.edu).

**COGENT (Cognitive Genetics Consortium)** — This work has been supported by grants from the National Institutes of Health (R01MH079800 and P50MH080173 to AKM; R01MH080912 to DCG; K23MH077807 to KEB; K01MH085812 to MCK). Data collection for the TOP cohort was supported by the Research Council of Norway, South-East Norway Health Authority, and KG Jebsen Foundation. The NCNG study was supported by Research Council of Norway Grants 154313/V50 and 177458/V50. The NCNG GWAS was financed by grants from the Bergen Research Foundation, the University of Bergen, the Research Council of Norway (FUGE, Psykisk Helse), Helse Vest RHF and Dr Einar Martens Fund. The Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland, the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, Signe and Ane Gyllenberg Foundation, University of Helsinki, Ministry of Education, Ahokas Foundation, Emil Aaltonen Foundation. For the LBC1936 cohort, phenotype collection was supported by The Disconnected Mind project. Genotyping was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC grant No. BB/F019394/1). The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative, which is funded by the Medical Research Council and the Biotechnology and Biological Sciences Research Council (MR/K026992/1). The CAMH work was supported by the CAMH Foundation and the Canadian Institutes of Health Research. The Duke Cognition Cohort (DCC) acknowledges K. Linney, J.M. McEvoy, P. Hunt, V. Dixon, T. Pennuto, K. Cornett, D. Swilling, L. Phillips, M. Silver, J. Covington, N. Walley, J. Dawson, H. Onabanjo, P. Nicoletti, A. Wagoner, J. Elmore, L. Bevan, J. Hunkin and R. Wilson for recruitment and testing of subjects. DCC also acknowledges the Ellison Medical Foundation New Scholar award AG-NS-0441-08 for partial funding of this study as well as the National Institute of Mental Health of the National Institutes of Health under award number K01MH098126. The UCLA Consortium for Neuropsychiatric Phenomics (CNP) study acknowledges the following sources of funding from the NIH: Grants UL1DE019580 and PL1MH083271 (RMB), RL1MH083269 (TDC), RL1DA024853 (EL) and PL1NS062410. The ASPIS study was supported by National Institute of Mental Health research grants R01MH085018 and R01MH092515 to Dr. Dimitrios Avramopoulos.

Support for the Duke Neurogenetics Study was provided the National Institutes of Health (R01 DA033369 and R01 AG049789 to ARH) and by a National Science Foundation Graduate Research Fellowship to MAS. Recruitment, genotyping and analysis of the TCD healthy control samples were supported by Science Foundation Ireland (grants 12/IP/1670, 12/IP/1359 and 08/IN.1/B1916). Data access for several cohorts used in this study was provided by the National Center for Biotechnology Information (NCBI) database of Genotypes and Phenotypes (dbGaP). dbGaP accession numbers for these cohorts were: Cardiovascular Health Study (CHS): phs000287.v4.p1, phs000377.v5.p1, and phs000226.v3.p1; Framingham Heart Study (FHS): phs000007.v23.p8 and phs000342.v11.p8; Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GENADA): phs000219.v1.p1; Long Life Family Study (LLFS): phs000397.v1.p1; Genetics of Late Onset Alzheimer's Disease Study (LOAD): phs000168.v1.p1; Minnesota Center for Twin and Family Research (MCTFR): phs000620.v1.p1; Philadelphia Neurodevelopmental Cohort (PNC): phs000607.v1.p1.

The acknowledgment statements for the cohorts included as part of the COGENT consortium are found below:

*Framingham Heart Study* — The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195 and HHSN268201500001I). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI. Funding for SHARe Affymetrix genotyping was provided by NHLBI Contract N02-HL-64278. SHARe Illumina genotyping was provided under an agreement between Illumina and Boston University.

*Cardiovascular Health Study* — This research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85084, N01HC85085, N01HC85086, N01HC35129, N01HC15103, N01HC55222, N01HC75150, N01HC45133, and N01HC85239; grant numbers U01HL080295 and U01HL130014 from the National Heart, Lung, and Blood Institute, and R01AG023629 from the National Institute on Aging, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal CHS investigators and institutions can be found at <https://chs-nhlbi.org/pi>. This manuscript was not prepared in collaboration with CHS investigators and does not necessarily reflect the opinions or views of CHS, or the NHLBI. Support for the genotyping through the CARE Study was provided by NHLBI Contract N01HC65226. Support for the Cardiovascular Health Study Whole Genome Study was provided by NHLBI grant HL087652. Additional

support for infrastructure was provided by HL105756 and additional genotyping among the African-American cohort was supported in part by HL085251, DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by National Center for Research Resources grant UL1RR033176, now at the National Center for Advancing Translational Technologies CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

*Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease* — The genotypic and associated phenotypic data used in the study were provided by the GlaxoSmithKline, R&D Limited. Details on data acquisition have been published previously in: Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, Hosford D, Barnes MR, Briley JD, Borrie M, Coletta N, Delisle R, Dhalla D, Ehm MG, Feldman HH, Fornazzari L, Gauthier S, Goodgame N, Guzman D, Hammond S, Hollingworth P, Hsiung GY, Johnson J, Kelly DD, Keren R, Kertesz A, King KS, Lovestone S, Loy-English I, Matthews PM, Owen MJ, Plumpton M, Pryse-Phillips W, Prinjha RK, Richardson JC, Saunders A, Slater AJ, St George-Hyslop PH, Stinnett SW, Swartz JE, Taylor RL, Wherrett J, Williams J, Yarnall DP, Gibson RA, Irizarry MC, Middleton LT, Roses AD. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch Neurol.*, Jan;65(1):45-53, 2008 (PMID: 17998437). Filippini N, Rao A, Wetten S, Gibson RA, Borrie M, Guzman D, Kertesz A, Loy-English I, Williams J, Nichols T, Whitcher B, Matthews PM. Anatomically-distinct genetic associations of APOE epsilon4 allele load with regional cortical atrophy in Alzheimer's disease. *Neuroimage*, Feb 1;44(3):724-8, 2009. (PMID: 19013250).

*Genetics of Late Onset Alzheimer's Disease Study* — Funding support for the "Genetic Consortium for Late Onset Alzheimer's Disease" was provided through the Division of Neuroscience, NIA. The Genetic Consortium for Late Onset Alzheimer's Disease includes a genome-wide association study funded as part of the Division of Neuroscience, NIA. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by Genetic Consortium for Late Onset Alzheimer's Disease. A list of contributing investigators is available at: [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000168.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000168.v1.p1)

*Long Life Family Study* — Funding support for the Long Life Family Study was provided by the Division of Geriatrics and Clinical Gerontology, National Institute on Aging. The Long Life Family Study includes GWAS analyses for factors that contribute to long and healthy life. Assistance with phenotype harmonization and genotype cleaning as well as with general study coordination, was provided by the Division of Geriatrics and Clinical Gerontology, National Institute on Aging. Support for the collection of

datasets and samples were provided by Multicenter Cooperative Agreement support by the Division of Geriatrics and Clinical Gerontology, National Institute on Aging (U01AG023746; U01AG023755; U01AG023749; U01AG023744; U01AG023712). Funding support for the genotyping which was performed at the Johns Hopkins University Center for Inherited Disease Research was provided by the National Institute on Aging, National Institutes of Health.

*Minnesota Center for Twin and Family Research* — This project was led by William G. Iacono, PhD. And Matthew K. McGue, PhD (Co-Principal Investigators) at the University of Minnesota, Minneapolis, MN, USA. Co-investigators from the same institution included: Irene J. Elkins, Margaret A. Keyes, Lisa N. Legrand, Stephen M. Malone, William S. Oetting, Michael B. Miller, and Saonli Basu. Funding support for this project was provided through NIDA (U01DA024417). Other support for sample ascertainment and data collection came from several grants: R37DA05147, R01AA09367, R01AA11886, R01DA13240, R01MH66140.

*Philadelphia Neurodevelopmental Cohort* — Support for the collection of the data sets was provided by grant RC2MH089983 awarded to Raquel Gur, MD, and RC2MH089924 awarded to Hakon Hakonarson, MD, PhD. All subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia.

**EGCUT4 (Estonian Genome Center)** — EGCUT received funding from the Estonian Research Council Grant IUT20-60 and PUT1660, EU H2020 grant 692145, and European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) GENTRANSMED. For more information, please contact Tõnu Esko ([tonu.esko@ut.ee](mailto:tonu.esko@ut.ee)).

**ELSA (English Longitudinal Study of Ageing)** — The English Longitudinal Study of Ageing is jointly run by University College London, Institute for Fiscal Studies, University of Manchester and National Centre for Social Research. Genetic analyses have been carried out by UCL Genomics and funded by the Economic and Social Research Council (ES/K005774/1) and the National Institute on Aging. All GWAS data has been deposited in the European Genome-phenome Archive. For more information contact [ELSAdata@natcen.ac.uk](mailto:ELSAdata@natcen.ac.uk).

**Fenland (Fenland Study)** — We are grateful to all the volunteers for their time and help and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Anthropometry, Data and Laboratory teams. The Fenland Study is supported by the Medical Research Council (MC\_U106179471). This work was supported by the Medical Research Council (Unit Programme numbers:



MC\_UU\_12015/1 and MC\_UU\_12015/2). Genotyping was supported by the Medical Research Council (MC\_PC\_13046). For inquiries about access to this data, please contact John Perry ([John.Perry@mrc-epid.cam.ac.uk](mailto:John.Perry@mrc-epid.cam.ac.uk)).

**Geisinger** — This work was supported by Geisinger Health System funds.

**GSII (Generation Scotland)** — We would like to acknowledge the contributions of the families who took part in the Generation Scotland: Scottish Family Health Study, the general practitioners and Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes academic researchers, IT staff, laboratory technicians, statisticians and research managers. Generation Scotland received core funding from the Chief Scientist Office of the Scottish Government Health Directorate CZD/16/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STRatifying Resilience & Depression Longitudinally (STRADL) Reference 104036/Z/14/Z). Information on applications for access to Generation Scotland data can be found at <http://www.generationscotland.org/>.

**Norfolk (EPIC- Norfolk Prospective Population Study)** — The authors would like to acknowledge the contribution of the staff and participants of the EPIC-Norfolk Study. EPIC-Norfolk is supported by the Medical Research Council (programme grants G0401527, G1000143) and Cancer Research UK (programme grant C864/A8257). This work was supported by the Medical Research Council (Unit Programme numbers MC\_UU\_12015/1 and MC\_UU\_12015/2). For inquiries about access to this data, please contact Ken Ong ([Ken.Ong@mrc-epid.cam.ac.uk](mailto:Ken.Ong@mrc-epid.cam.ac.uk)).

**UKHLS (The UK Household Longitudinal Study)** — Understanding Society: The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council (Grant Number: ES/M008592/1). Data were collected by NatCen and the genome wide scan data were analysed by the Wellcome Trust Sanger Institute. Information on how to access the data can be found at <https://www.understandingsociety.ac.uk/>.

**VIKING (Viking Health Study – Shetland)** — The Viking Health Study- Shetland (VIKING) was supported by the MRC Human Genetics Unit quinquennial programme grant QTL in Health and Disease. DNA extractions and genotyping were performed at the Edinburgh Clinical Research Facility, University of Edinburgh. We would like to acknowledge the invaluable contributions of the research nurses in Shetland, the

administrative team in Edinburgh and the people of Shetland. For inquiries about access to this data, please contact Jim Wilson ([jim.wilson@ed.ac.uk](mailto:jim.wilson@ed.ac.uk)).

**WLS (Wisconsin Longitudinal Study)** — This research uses data from the Wisconsin Longitudinal Study (*WLS*) of the University of Wisconsin-Madison. Since 1991, the *WLS* has been supported principally by the National Institute on Aging (AG-9775, AG-21079, AG-033285, and AG-041868), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. Since 1992, data have been collected by the University of Wisconsin Survey Center. The opinions expressed herein are those of the authors. A public use file of data from the Wisconsin Longitudinal Study is available from the Wisconsin Longitudinal Study, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at <http://www.ssc.wisc.edu/WLSresearch/data/>.

**UK Biobank** — This research has also been conducted using the UK Biobank Resource under Application Numbers 11425 and 12514. Informed consent was obtained from UK Biobank subjects.

**Individual Acknowledgements** — Peter M. Visscher gratefully acknowledges funding from the Australian National Health and Medical Research Council (1080157, 1078037, 1048853, 1050218, and 1113400). Jian Yang also acknowledges funding from the Australian National Health and Medical Research Council (1107258 and 1113400), from the Australian Research Council (Discovery Project 160101343), and from a Charles and Sylvia Viertel Foundation Senior Medical Research Fellowship. Matthew R. Robinson acknowledges funding from the Australian Research Council (Discovery Project 160103860). Tune H. Pers and Pascal N. Timshel acknowledge funding from the Lundbeck Foundation. Philipp D. Koellinger, Aysu Okbay, and Richard Karlsson Linnér acknowledge funding from an ERC Consolidator Grant to Philipp D. Koellinger (647648 EdGe). Dalton Conley acknowledges funding from Russell Sage Foundation Grant #83-15-29: GxE and Health Inequality across the Life Course. Robbee Wedow was generously supported by the National Science Foundation's Graduate Research Fellowship Program (DGE 1144083). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## 7.4 Extended Acknowledgements

This paper uses cohort level data from Okbay et al.<sup>1</sup>. Information about studies participating in that study can be found in the Additional Acknowledgements Supplemental chapter of that paper. Per SSGAC policy, below we acknowledge the authors of that paper:

Aysu Okbay, Jonathan P. Beauchamp, Mark Alan Fontana, James J. Lee, Tune H. Pers, Cornelius A. Rietveld, Patrick Turley, Guo-Bo Chen, Valur Emilsson, S. Fleur W. Meddens, Sven Oskarsson, Joseph K. Pickrell, Kevin Thom, Pascal Timshel, Ronald de Vlaming, Abdel Abdellaoui, Tarunveer S. Ahluwalia, Jonas Bacelis, Clemens Baumbach, Gyda Bjornsdottir, Johannes H. Brandsma, Maria Pina Concas, Jaime Derringer, Nicholas A. Furlotte, Tessel E. Galesloot, Giorgia Grotto, Richa Gupta, Leanne M. Hall, Sarah E. Harris, Edith Hofer, Momoko Horikoshi, Jennifer E. Huffman, Kadri Kaasik, Ioanna P. Kalafati, Robert Karlsson, Augustine Kong, Jari Lahti, Sven J. van der Lee, Christiaan deLeeuw, Penelope A. Lind, Karl-Oskar Lindgren, Tian Liu, Massimo Mangino, Jonathan Marten, Evelin Mihailov, Michael B. Miller, Peter J. van der Most, Christopher Oldmeadow, Antony Payton, Natalia Pervjakova, Wouter J. Peyrot, Yong Qian, Olli Raitakari, Rico Rueedi, Erika Salvi, Börge Schmidt, Katharina E. Schraut, Jianxin Shi, Albert V. Smith, Raymond A. Poot, Beate St Pourcain, Alexander Teumer, Gudmar Thorleifsson, Niek Verweij, Dragana Vuckovic, Juergen Wellmann, Harm-Jan Westra, Jingyun Yang, Wei Zhao, Zhihong Zhu, Behrooz Z. Alizadeh, Najaf Amin, Andrew Bakshi, Sebastian E. Baumeister, Ginevra Biino, Klaus Bønnelykke, Patricia A. Boyle, Harry Campbell, Francesco P. Cappuccio, Gail Davies, Jan-Emmanuel De Neve, Panos Deloukas, Ilja Demuth, Jun Ding, Peter Eibich, Lewin Eisele, Niina Eklund, David M. Evans, Jessica D. Faul, Mary F. Feitosa, Andreas J. Forstner, Ilaria Gandin, Bjarni Gunnarsson, Bjarni V. Halldórsson, Tamara B. Harris, Andrew C. Heath, Lynne J. Hocking, Elizabeth G. Holliday, Georg Homuth, Michael A. Horan, Jouke-Jan Hottenga, Philip L. de Jager, Peter K. Joshi, Astanand Jugessur, Marika A. Kaakinen, Mika Kähönen, Stavroula Kanoni, Liisa Keltigangas-Järvinen, Lambertus A. L. M. Kiemeney, Ivana Kolcic, Seppo Koskinen, Aldi T. Kraja, Martin Kroh, Zoltan Kutalik, Antti Latvala, Lenore J. Launer, Maël P. Lebreton, Douglas F. Levinson, Paul Lichtenstein, Peter Lichtner, David C. M. Liewald, LifeLines Cohort Study, Anu Loukola, Pamela A. Madden, Reedik Mägi, Tomi Mäki-Opas, Riccardo E. Marioni, Pedro Marques-Vidal, Gerardus A. Meddens, George McMahon, Christa Meisinger, Thomas Meitinger, Yusplitri Milaneschi, Lili Milani, Grant W. Montgomery, Ronny Myhre, Christopher P. Nelson, Dale R. Nyholt, William E. R. Ollier, Aarno Palotie, Lavinia Paternoster, Nancy L. Pedersen, Katja E. Petrovic, David J. Porteous, Katri Räikkönen, Susan M. Ring, Antonietta Robino, Olga Rostapshova, Igor Rudan, Aldo Rustichini, Veikko Salomaa,

Alan R. Sanders, Antti-Pekka Sarin, Helena Schmidt, Rodney J. Scott, Blair H. Smith, Jennifer A. Smith, Jan A. Staessen, Elisabeth Steinhagen-Thiessen, Konstantin Strauch, Antonio Terracciano, Martin D. Tobin, Sheila Ulivi, Simona Vaccargiu, Lydia Quayle, Frank J. A. van Rooij, Cristina Venturini, Anna A. E. Vinkhuyzen, Uwe Völker, Henry Völzke, Judith M. Vonk, Diego Vozzi, Johannes Waage, Erin B. Ware, Gonneke Willemsen, John R. Attia, David A. Bennett, Klaus Berger, Lars Bertram, Hans Bisgaard, Dorret I. Boomsma, Ingrid B. Borecki, Ute Bültmann, Christopher F. Chabris, Francesco Cucca, Daniele Cusi, Ian J. Deary, George V. Dedoussis, Cornelia M. van Duijn, Johan G. Eriksson, Barbara Franke, Lude Franke, Paolo Gasparini, Pablo V. Gejman, Christian Gieger, Hans-Jörgen Grabe, Jacob Gratten, Patrick J. F. Groenen, Vilmundur Gudnason, Pim van der Harst, Caroline Hayward, David A. Hinds, Wolfgang Hoffmann, Elina Hyppönen, William G. Iacono, Bo Jacobsson, Marjo-Riitta Järvelin, Karl-Heinz Jöckel, Jaakko Kaprio, Sharon L. R. Kardia, Terho Lehtimäki, Steven F. Lehrer, Patrik K. E. Magnusson, Nicholas G. Martin, Matt McGue, Andres Metspalu, Neil Pendleton, Brenda W. J. H. Penninx, Markus Perola, Nicola Pirastu, Mario Pirastu, Ozren Polasek, Danielle Posthuma, Christine Power, Michael A. Province, Nilesh J. Samani, David Schlessinger, Reinhold Schmidt, Thorkild I. A. Sørensen, Tim D. Spector, Kari Stefansson, Unnur Thorsteinsdottir, A. Roy Thurik, Nicholas J. Timpson, Henning Tiemeier, Joyce Y. Tung, André G. Uitterlinden, Veronique Vitart, Peter Vollenweider, David R. Weir, James F. Wilson, Alan F. Wright, Dalton C. Conley, Robert F. Krueger, George Davey Smith, Albert Hofman, David I. Laibson, Sarah E. Medland, Michelle N. Meyer, Jian Yang, Magnus Johannesson, Peter M. Visscher, Tõnu Esko, Philipp D. Koellinger, David Cesarini & Daniel J. Benjamin.

We gratefully acknowledge the contributions of members of 23andMe's Research Team, whose names are listed below:

Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Catherine H. Wilson, and Steven J. Pitts.

We also gratefully acknowledge the contributions of members of the Cognitive Genomics Consortium (COGENT), whose names are listed below:

Ole A. Andreassen, Dan E. Arking, Deborah K. Attix, Dimitrios Avramopoulos, Robert M. Bilder, Panos Bitsios, Katherine E. Burdick, Tyrone D. Cannon, Ornit Chiba-Falek, Andrea Christoforou, Elizabeth T. Cirulli, Eliza Congdon,

Emily Drabant Conley, Aiden Corvin, Gail Davies, Ian J. Deary, Pamela DeRosse, Dwight Dickinson, Srdjan Djurovic, Gary Donohoe, Johan G. Eriksson, Thomas Espeseth, Nelson A. Freimer, Stella Giakoumaki, Ina Giegling, Michael Gill, David C. Glahn, Ahmad R. Hariri, Alex Hatzimanolis, Matthew C. Keller, Emma Knowles, Bettina Konte, Jari Lahti, Stephanie Le Hellard, David C. Liewald, Edythe London, Astri J. Lundervold, Ingrid Melle, Derek Morris, Anna C. Need, William Ollier, Aarno Palotie, Antony Payton, Neil Pendleton, Russell A. Poldrack, Katri Räikkönen, Ivar Reinvang, Panos Roussos, Dan Rujescu, Fred W. Sabb, Matthew A. Scult, Nikolaos Smyrnis, John M. Starr, Vidar M. Steen, Nikos C. Stefanis, Richard E. Straub, Kjetil Sundet, Aristotle N. Voineskos, Daniel R. Weinberger, Elisabeth Widen, Jin Yu.

## 8. References

1. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
2. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
3. The 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
4. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
5. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
6. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
7. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
8. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
9. Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10**, 101 (1954).
10. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
11. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
12. Trampush, J. *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Nat. Publ. Gr.* (2017). doi:10.1038/mp.2016.244
13. Turley, P. *et al.* MTAG: Multi-Trait Analysis of GWAS. *Nat. Genet.* **in press**, (2017).
14. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
15. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3**, e3395 (2008).
16. Rietveld, Cornelius; Conley, Dalton; Eriksson, Nicholas; Koellinger, P. Replicability and robustness of GWAS for behavioral traits. *Psychol. Sci.* (2014).
17. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16 (2015).
18. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
19. Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nat. Hum. Behav.* (2017). doi:10.1038/s41562-016-0016
20. Fisher, R. A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh* **52**, 399–433 (1918).
21. Fisher, R. A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh* **52**, 399–433 (1918).

22. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
23. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
24. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
25. Branigan, A. R. *et al.* Variation in the Heritability of Educational Attainment: An International Meta-Analysis. *Soc. Forces* **92**, 109–140 (2013).
26. Sacerdote, B. in *Handbook of Social Economics* (eds. Benhabib, J., Bisin, A. & Jackson, M. O.) 1–29 (Elsevier/North-Holland, 2011).
27. Sacerdote, B. How Large are the Effects from Changes in Family Environment? A Study of Korean American Adoptees. *Q. J. Econ.* **122**, 119–157 (2007).
28. Björklund, A., Lindahl, M. & Plug, E. The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data\*. *Q. J. Econ.* **121**, 999–1028 (2006).
29. Cesarini, D. & Visscher, P. M. Genetics and educational attainment. *npj Sci. Learn.* **2**, 4 (2017).
30. Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M. & Boardman, J. D. Polygenic Influence on Educational Attainment: New evidence from The National Longitudinal Study of Adolescent to Adult Health. *AERA Open* **1**, 1–13 (2015).
31. Belsky, D. W. *et al.* The Genetics of Success. *Psychol. Sci.* **27**, 957–972 (2016).
32. Liang, K.-Y. & Zeger, S. L. Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22 (1986).
33. Lee, J. C. Correlation and causation in the study of personality (with discussion). *Eur. J. Pers.* **26**, 372–412 (2012).
34. Crow, J. F. & Kimura, M. *An Introduction to Population Genetics Theory*. (Harper & Row, 1970). doi:10.2307/1529706
35. Lee, J. C. & Chow, C. C. LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *bioRxiv* (2017).
36. de Vlaming, R. *et al.* Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *bioRxiv* (2016). doi:10.1101/048322
37. Cameron, A. C. & Miller, D. Robust inference with dyadic data. *mimeo* (2014). doi:10.1201/b10440
38. Barro, R. J. & Lee, J. W. A new data set of educational attainment in the world, 1950–2010. *J. Dev. Econ.* **104**, 184–198 (2013).
39. Su, L. & Ullah, A. Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing. *Handb. Empir. Econ. Financ.* 455–497 (2011). doi:10.1201/b10440
40. Lynch, M. & Walsh, B. in *Genetics and Analysis of Quantitative Traits* 980 (Sinauer Associates, Inc., 1998). doi:10.1086/318209
41. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).
42. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017). doi:http://dx.doi.org/10.1101/166298
43. Goddard, M. E., Hayes, B. J. & Meuwissen, T. H. E. Using the genomic

- relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* **128**, 409–421 (2011).
44. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* **17**, 9–18 (2016).
  45. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
  46. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat. Genet.* **47**, 115–125 (2015).
  47. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976 (2007).
  48. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48**, 856–866 (2016).
  49. Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
  50. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* (2017).
  51. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
  52. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *bioRxiv* (2017). doi:10.1101/103069
  53. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
  54. de Leeuw, C. A. *et al.* MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
  55. Liu, J. Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).
  56. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).
  57. Mi, H. *et al.* PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45**, D183–D189 (2017).
  58. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).
  59. Gazal, S. *et al.* Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
  60. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
  61. Chen, W. *et al.* Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* **200**, 719–736 (2015).
  62. Chen, W., McDonnell, S. K., Thibodeau, S. N., Tillmans, L. S. & Schaid, D. J. Incorporating functional annotations for fine-mapping causal variants in a Bayesian framework using summary statistics. *Genetics* **204**, 933–958 (2016).
  63. Hastie, T., Tibshirani, R. & Wainwright, M. *Statistical Learning with Sparsity:*



- The Lasso and Generalizations*. (CRC Press, 2015).
64. Gazal, S. *et al.* OLD Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature Genetics* (2017). doi:10.1038/ng.3954
  65. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).
  66. Deary, I. J. *Looking Down on Human Intelligence: From Psychometrics to the Brain*. (Oxford University Press, 2000).
  67. Jensen, A. R. *Clocking the Mind: Mental Chronometry and Individual Differences*. (Elsevier, 2006).
  68. Ratcliff, R., Schmiedek, F. & McKoon, G. A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence* **36**, 10–17 (2008).
  69. van Ravenzwaaij, D., Brown, S. & Wagenmakers, E.-J. An integrated perspective on the relation between response speed and intelligence. *Cognition* **119**, 381–393 (2011).
  70. Ronan, J. L., Wu, W. & Crabtree, G. R. From neural development to cognition: Unexpected roles for chromatin. *Nat. Rev. Genet.* **14**, 347–359 (2013).
  71. Hota, S. K. & Bruneau, B. G. ATP-dependent chromatin remodeling during mammalian development. *Development* **143**, 2882–2897 (2016).
  72. Lessard, J. *et al.* An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201–215 (2007).
  73. Son, E. Y. & Crabtree, G. R. The role of BAF (mSWI/SNF) complexes in mammalian neural development. *Am. J. Med. Genet. Part C* **166C**, 333–349 (2014).
  74. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).
  75. Tuoc, T. C. *et al.* Chromatin regulation by BAF170 controls cerebral cortical size and thickness. *Dev. Cell* **25**, 256–269 (2013).
  76. Nagamani, S. C. S. *et al.* Interstitial deletion of 6q25.2-q25.3: A novel microdeletion syndrome associated with microcephaly, developmental delay, dysmorphic features and hearing loss. *Eur. J. Hum. Genet.* **17**, 573–581 (2009).
  77. Hoyer, J. *et al.* Haploinsufficiency of ARID1B, a member of the SWI/SNF-A chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.* **90**, 565–572 (2012).
  78. Vals, M.-A. *et al.* Coffin-Siris Syndrome with obesity, macrocephaly, hepatomegaly and hyperinsulinism caused by a mutation in the ARID1B gene. *Eur. J. Hum. Genet.* **22**, 1327–1329 (2014).
  79. Van Houdt, J. K. J. *et al.* Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat. Genet.* **44**, 445–449 (2012).
  80. Ninkovic, J. *et al.* The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. *Cell Stem Cell* **13**, 403–418 (2013).
  81. Tsurusaki, Y. *et al.* De novo SOX11 mutations cause Coffin-Siris syndrome. *Nat. Commun.* **5**, 4011 (2014).
  82. Hempel, A. *et al.* Deletions and de novo mutl. A. Hempel et al., Deletions and de

- novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of Coffin-Siris syndrome. *J. Med. Genet.* **53**, 152–162 (2016).ations of SOX11 are associated with a neurod. *J. Med. Genet.* **53**, 152–162 (2016).
83. de Dieuleveult, M. *et al.* Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. *Nature* **530**, 113–116 (2016).
  84. Carvill, G. L. *et al.* Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat. Genet.* **45**, 825–830 (2013).
  85. Suls, A. *et al.* De novo loss-of-function mutations in CHD2 cause a fever-sensitive myoclonic epileptic encephalopathy sharing features with Dravet syndrome. *Am. J. Hum. Genet.* **93**, 967–975 (2013).
  86. Shen, T., Ji, F., Yuan, Z. & Jiao, J. CHD2 is required for embryonic neurogenesis in the developing cerebral cortex. *Stem Cells* **33**, 1794–1806 (2015).
  87. Krumm, N., O’Roak, B. J., Shendure, J. & Eichler, E. E. A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci.* **37**, 95–105 (2014).
  88. Robinson, E. B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).
  89. Williams, C. A., Dagli, A. & Battaglia, A. Genetic disorders associated with macrocephaly. *Am. J. Med. Genet. Part A* **146A**, 2023–2037 (2008).
  90. Hazlett, H. C. *et al.* Early brain development in infants at high risk for autism spectrum disorder. *Nature* **542**, 348–351 (2017).
  91. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. USA* **111**, E4468–E4477 (2014).
  92. Katayama, Y. *et al.* CHD8 haploinsufficiency results in autistic-like phenotypes in mice. *Nature* **537**, 675–679 (2016).
  93. Torchy, M. P., Hamiche, A. & Klaholz, B. P. Structure and function insights into the NuRD chromatin remodeling complex. *Cell. Mol. Life Sci.* **72**, 2491–2507 (2015).
  94. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
  95. Willemsen, M. H. *et al.* GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. *J. Med. Genet.* **50**, 507–514 (2013).
  96. Kaidi, A. & Jackson, S. P. KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. *Nature* **498**, 70–74 (2013).
  97. Penicud, K. & Behrens, A. DMAP1 is an essential regulator of ATM activity and function. *Oncogene* **33**, 525–531 (2014).
  98. Huang, J., Gong, Z., Ghosal, G. & Chen, J. SOSS complexes participate in the maintenance of genomic stability. *Mol. Cell* **35**, 384–393 (2009).
  99. Clayton-Smith, J. *et al.* Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. *Am. J. Hum. Genet.* **89**, 675–681 (2011).
  100. Campeau, P. M. *et al.* Mutations in KAT6B, encoding a histone acetyltransferase,

- cause Genitopatellar syndrome. *Am. J. Hum. Genet.* **90**, 282–289 (2012).
101. Simpson, M. A. *et al.* De novo mutations of the gene encoding the histone acetyltransferase KAT6B cause Genitopatellar syndrome. *Am. J. Hum. Genet.* **90**, 290–294 (2012).
  102. Weake, V. M. & Workman, J. L. SAGA function in tissue-specific gene expression. *Trends Cell Biol.* **22**, 177–184 (2012).
  103. Bonnet, J. *et al.* The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev.* **28**, 1999–2012 (2014).
  104. Morgan, M. T. *et al.* Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**, 725–728 (2016).
  105. Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* **493**, 699–702 (2013).
  106. Tan, S.-L. *et al.* Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development* **139**, 3806–3816 (2012).
  107. Fei, Q. *et al.* SETDB1 modulates PRC2 activity at developmental genes independently of H3K9 trimethylation in mouse ES cells. *Genome Res.* **25**, 1325–1335 (2015).
  108. Beck, D. B., Oda, H., Shen, S. S. & Reinberg, D. PR-Set7 and H4K20me1: At the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev.* **26**, 325–337 (2012).
  109. Lan, F. *et al.* Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**, 718–722 (2007).
  110. Kim, H.-G. *et al.* Translocations disrupting PHF21A in the Potocki-Shaffer-syndrome region are associated with intellectual disability and craniofacial anomalies. *Am. J. Hum. Genet.* **91**, 56–72 (2012).
  111. Montgomery, N. D., Turcott, C. M., Tepperberg, J. H., McDonald, M. T. & Aylsworth, A. S. A 137-kb deletion within the Potocki-Shaffer syndrome interval on chromosome 11p11.2 associated with developmental delay and hypotonia. *Am. J. Med. Genet. Part A* **161A**, 198–202 (2013).
  112. Labonne, J. D. J. *et al.* A microdeletion encompassing PHF21A in an individual with global developmental delay and craniofacial anomalies. *Am. J. Med. Genet. Part A* **167**, 3011–3018 (2015).
  113. Dopie, J., Skarp, K.-P., Rajakylä, E. K., Tanhuanpää, K. & Vartiainen, M. K. Active maintenance of nuclear actin by importin 9 supports transcription. *Proc. Natl. Acad. Sci. USA* **109**, E544–552 (2012).
  114. Nishimoto, S., Kusakabe, M. & Nishida, E. Requirement of the MEK5-ERK5 pathway for neural differentiation in *Xenopus* embryonic development. *EMBO Rep.* **6**, 1064–1069 (2005).
  115. Liu, L. *et al.* Extracellular signal-regulated kinase (ERK) 5 is necessary and sufficient to specify cortical neuronal fate. *Proc. Natl. Acad. Sci. USA* **103**, 9697–9702 (2006).
  116. Stankiewicz, T. R., Ramaswami, S. A., Bouchard, R. J., Aktories, K. & Linseman, D. A. Neuronal apoptosis induced by selective inhibition of Rac GTPase versus global suppression of Rho family GTPases is mediated by alterations in distinct

- mitogen-activated protein kinase signaling cascades. *J. Biol. Chem.* **290**, 9363–9376 (2015).
117. Impey, S., Obrietan, K. & Storm, D. R. Making new connections: Role of ERK/MAP kinase signaling in neuronal plasticity. *Neuron* **23**, 11–14 (1999).
  118. Krens, S. F. G., Spaink, H. P. & Snaar-Jagalska, B. E. Functions of the MAPK family in vertebrate-development. *FEBS Lett.* **580**, 4984–4990 (2006).
  119. Morgan, D. O. Principles of CDK regulation. *Nature* **374**, 131–134 (1995).
  120. Malumbres, M. Cyclin-dependent kinases. *Genome Biol.* **15**, 110–122 (2014).
  121. Blazek, D. *et al.* The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev.* **25**, 2158–2172 (2011).
  122. Mukhopadhyay, A. *et al.* CDK19 is disrupted in a female patient with bilateral congenital retinal folds, microcephaly and mild mental retardation. *Hum. Genet.* **128**, 281–291 (2010).
  123. Yin, J.-W. & Wang, G. The Mediator complex: A master coordinator of transcription and cell lineage development. *Development* **141**, 977–987 (2014).
  124. Adegbola, A. *et al.* Redefining the MED13L syndrome. *Eur. J. Hum. Genet.* **23**, 1308–1317 (2015).
  125. Cafiero, C. *et al.* Novel de novo heterozygous loss-of-function variants in MED13L and further delineation of the MED13L haploinsufficiency syndrome. *Eur. J. Hum. Genet.* **23**, 1499–1504 (2015).
  126. Ray, D. *et al.* A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
  127. Sutherland, L. C., Rintala-Maki, N. D., White, R. D. & Morin, C. D. RNA binding motif (RBM) proteins: A novel family of apoptosis modulators? *J. Cell. Biochem.* **94**, 5–24 (2005).
  128. Telley, L. *et al.* Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**, 1443–1446 (2016).
  129. Vuong, C. K., Black, D. L. & Zheng, S. The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016).
  130. Yin, X. *et al.* Dual-specificity tyrosine phosphorylation-regulated kinase 1A (Dyrk1A) modulates serine/arginine-rich protein 55 (SRp55)-promoted Tau exon 10 inclusion. *J. Biol. Chem.* **287**, 30497–30506 (2012).
  131. Fernández-Nogales, M. *et al.* Huntington’s disease is a four-repeat tauopathy with tau nuclear rods. *Nat. Med.* **20**, 881–885 (2014).
  132. Kamma, H., Portman, D. S. & Dreyfuss, G. Cell type-specific expression of hnRNP proteins. *Exp. Cell Res.* **221**, 187–196 (1995).
  133. Krecic, A. M. & Swanson, M. S. hnRNP complexes: Composition, structure, and function. *Curr. Opin. Cell Biol.* **11**, 363–371 (1999).
  134. Lin, N. *et al.* An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. *Mol. Cell* **53**, 1005–1019 (2014).
  135. Liu, Y. & Szaro, B. G. hnRNP K post-transcriptionally co-regulates multiple cytoskeletal genes needed for axonogenesis. *Development* **138**, 3079–3090 (2011).
  136. Yano, M., Okano, H. J. & Okano, H. Involvement of Hu and heterogeneous nuclear ribonucleoprotein K in neuronal differentiation through p21 mRNA post-transcriptional regulation. *J. Biol. Chem.* **280**, 12690–12699 (2005).

137. Scoumanne, A., Cho, S. J., Zhang, J. & Chen, X. The cyclin-dependent kinase inhibitor p21 is regulated by RNA-binding protein PCBP4 via mRNA stability. *Nucleic Acids Res.* **39**, 213–224 (2011).
138. Dowling, R. J. O. *et al.* mTORC1-mediated cell proliferation, but not cell growth, controlled by the 4E-BPs. *Science* **328**, 1172–1176 (2010).
139. Thoreen, C. C. *et al.* A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature* **485**, 109–113 (2012).
140. Zhou, J. *et al.* Pharmacological inhibition of mTORC1 suppresses anatomical, cellular, and behavioral abnormalities in neural-specific Pten knock-out mice. *J. Neurosci.* **29**, 1773–1783 (2009).
141. Palazuelos, J., Ortega, Z., Diaz-Alonso, J., Guzmán, M. & Galve-Roperh, I. CB2 cannabinoid receptors promote neural progenitor cell proliferation via mTORC1 signaling. *J. Biol. Chem.* **287**, 1198–1209 (2012).
142. Cloëtta, D. *et al.* Inactivation of mTORC1 in the developing brain causes microcephaly and affects gliogenesis. *J. Neurosci.* **33**, 7799–7810 (2013).
143. Hartman, N. W. *et al.* mTORC1 targets the translational repressor 4E-BP2, but not S6 kinase 1/2, to regulate neural stem cell self-renewal in vivo. *Cell Rep.* **5**, 433–444 (2013).
144. Bukhari, S. I. A. *et al.* A specialized mechanism of translation mediated by FXR1a-associated microRNP in cellular quiescence. *Mol. Cell* **61**, 760–773 (2016).
145. Nishi, K., Nishi, A., Nagasawa, T. & Ui-Tei, K. Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus. *RNA* **19**, 17–35 (2013).
146. Katz, S. *et al.* A nuclear role for miR-9 and Argonaute proteins in balancing quiescent and activated neural stem cell states. *Cell Rep.* **17**, 1383–1398 (2016).
147. Gehman, L. T. *et al.* The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).
148. Lee, J.-A. *et al.* Cytoplasmic Rbfox1 regulates the expression of synaptic and autism-related genes. *Neuron* **89**, 113–128 (2016).
149. Adams, H. H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, 1569–1582 (2016).
150. Manning, B. D. & Cantley, L. C. AKT/PKB signaling: Navigating downstream. *Cell* **129**, 1261–1274 (2007).
151. Hemmings, B. A. & Restuccia, D. F. PI3K-PKB/AKT pathway. *Cold Spring Harb. Perspect. Biol.* **4**, a011189 (2012).
152. Martini, M., De Santis, M. C., Braccini, L., Gulluni, F. & Hirsch, E. PI3K/AKT signaling pathway and cancer: An updated review. *Ann. Med.* **46**, 372–383 (2014).
153. Brunet, A., Datta, S. R. & Greenberg, M. E. Transcription-dependent and -independent control of neuronal survival by the PI3K-AKT signaling pathway. *Curr. Opin. Neurobiol.* **11**, 297–305 (2001).
154. Zhang, Y. E. Non-Smad pathways in TGF- $\beta$  signaling. *Cell Res.* **19**, 128–139 (2009).
155. Lehtinen, M. K. & Walsh, C. A. Neurogenesis at the brain-cerebrospinal fluid interface. *Annu. Rev. Cell Dev. Biol.* **27**, 653–679 (2011).
156. Shen, Q. *et al.* Endothelial cells stimulate self-renewal and expand neurogenesis of

- neural stem cells. *Science* **304**, 1338–1340 (2004).
157. Yuzwa, S. A. *et al.* Proneurogenic ligands defined by modeling developing cortex growth factor communication networks. *Neuron* **91**, 988–1004 (2016).
  158. Laplante, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274–293 (2012).
  159. Feng, J., Park, J., Cron, P., Hess, D. & Hemmings, B. A. Identification of a PKB/AKT hydrophobic motif Ser-473 kinase as DNA-dependent protein kinase. *J. Biol. Chem.* **279**, 41189–41196 (2004).
  160. Lopez-Pajares, V., Kim, M. M. & Yuan, Z.-M. Phosphorylation of MDMX mediated by AKT leads to stabilization and induces 14-3-3 binding. *J. Biol. Chem.* **283**, 13707–13713 (2008).
  161. Perry, M. E. The regulation of the p53-mediated stress response by MDM2 and MDM4. *Cold Spring Harb. Perspect. Biol.* **2**, a000968 (2010).
  162. Wade, M., Wang, Y. V & Wahl, G. M. The p53 orchestra: Mdm2 and Mdmx set the tone. *Trends Cell Biol.* **20**, 299–309 (2010).
  163. Kuan, C.-Y., Roth, K. A., Flavell, R. A. & Rakic, P. Mechanisms of programmed cell death in the developing brain. *Trends Neurosci.* **23**, 291–297 (2000).
  164. Dansen, T. B. & Burgering, B. M. T. Unravelling the tumor-suppressive functions of FOXO proteins. *Trends Cell Biol.* **18**, 421–429 (2008).
  165. Jacobs, F. M. J. *et al.* FoxO6, a novel member of the FoxO class of transcription factors with distinct shuttling dynamics. *J. Biol. Chem.* **278**, 35959–35967 (2003).
  166. de la Torre-Ubieta, L. & Bonni, A. Transcriptional regulation of neuronal polarity and morphogenesis in the mammalian brain. *Neuron* **72**, 22–40 (2011).
  167. Mattson, M. P. & Meffert, M. K. Roles for NF- $\kappa$ B in nerve cell survival, plasticity, and disease. *Cell Death Differ.* **13**, 852–860 (2006).
  168. Iwashita, Y., Fukuchi, N., Waki, M., Hayashi, K. & Tahira, T. Genome-wide repression of NF- $\kappa$ B target genes by transcription factor MIBP1 and its modulation by O-linked  $\beta$ -N-acetylglucosamine (O-GlcNAc) transferase. *J. Biol. Chem.* **287**, 9887–9900 (2012).
  169. Sun, Y., Goderie, S. K. & Temple, S. Asymmetric distribution of EGFR receptor during mitosis generates diverse CNS progenitor cells. *Neuron* **45**, 873–886 (2005).
  170. Mirzaa, G. M. & Poduri, A. Megalencephaly and hemimegalencephaly: Breakthroughs in molecular etiology. *Am. J. Med. Genet. Part C* **166C**, 156–172 (2014).
  171. Anderson, M. F., Aberg, M. A. I., Nilsson, M. & Eriksson, P. S. Insulin-like growth factor-I and neurogenesis in the adult mammalian brain. *Dev. Brain Res.* **134**, 115–122 (2002).
  172. Easton, R. M. *et al.* Role for AKT3/protein kinase B $\gamma$  in attainment of normal brain size. *Mol. Cell. Biol.* **25**, 1869–1878 (2005).
  173. Boland, E. *et al.* Mapping of deletion and translocation breakpoints in 1q44 implicates the serine/threonine kinase AKT3 in postnatal microcephaly and agenesis of the corpus callosum. *Am. J. Hum. Genet.* **81**, 292–303 (2007).
  174. Rivière, J.-B. *et al.* De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat. Genet.* **44**, 934–940 (2012).

175. Poduri, A. *et al.* Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* **74**, 41–48 (2012).
176. Lee, J. H. *et al.* De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat. Genet.* **44**, 941–945 (2012).
177. Carson, R. P., Van Nielen, D. L., Winzenburger, P. A. & Ess, K. C. Neuronal and glia abnormalities in Tsc1-deficient forebrain and partial rescue by rapamycin. *Neurobiol. Dis.* **45**, 369–380 (2012).
178. Cecconi, F., Alvarez-Bolado, G., Meyer, B. I., Roth, K. A. & Gruss, P. Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
179. Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
180. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
181. Srivastava, S. *et al.* Loss-of-function variants in HIVEP2 are a cause of intellectual disability. *Eur. J. Hum. Genet.* **24**, 556–561 (2016).
182. Steinfeld, H. *et al.* Mutations in HIVEP2 are associated with developmental delay, intellectual disability, and dysmorphic features. *Neurogenetics* **17**, 159–164 (2016).
183. Neves, G. *et al.* The LIM homeodomain protein Lhx6 regulates maturation of interneurons and network excitability in the mammalian cortex. *Cereb. Cortex* **23**, 1811–1823 (2013).
184. Vogt, D. *et al.* Lhx6 directly regulates Arx and CXCR7 to determine cortical interneuron fate and laminar position. *Neuron* **82**, 350–364 (2014).
185. Tissir, F. & Goffinet, A. M. Reelin and brain development. *Nat. Rev. Neurosci.* **4**, 496–505 (2003).
186. Britto, J. M. *et al.* Exogenous Reelin modifies the migratory behavior of neurons depending on cortical location. *Cereb. Cortex* **24**, 2835–2847 (2014).
187. Gao, Z. & Godbout, R. Reelin-Disabled-1 signaling in neuronal migration: Splicing takes the stage. *Cell. Mol. Life Sci.* **70**, 2319–2329 (2013).
188. Bock, H. H. & May, P. Canonical and non-canonical Reelin signaling. *Front. Cell. Neurosci.* **10**, 166 (2016).
189. Takeichi, M. The cadherin superfamily in neuronal connections and interactions. *Nat. Rev. Neurosci.* **8**, 11–20 (2007).
190. Zhang, J. *et al.* Cortical neural precursors inhibit their own differentiation via N-cadherin maintenance of  $\beta$ -catenin signaling. *Dev. Cell* **18**, 472–479 (2010).
191. Matsunaga, Y. *et al.* Reelin transiently promotes N-cadherin-dependent neuronal adhesion during mouse cortical development. *Proc. Natl. Acad. Sci. USA* **114**, 2048–2053 (2017).
192. Moon, U. Y. *et al.* Impaired Reelin-Dab1 signaling contributes to neuronal migration deficits of tuberous sclerosis complex. *Cell Rep.* **12**, 965–978 (2015).
193. Hong, S. E. *et al.* Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN mutations. *Nat. Genet.* **26**, 93–96 (2000).
194. Zhang, X. *et al.* SUN1/2 and Syne/Nesprin-1/2 complexes connect centrosome to the nucleus during neurogenesis and neuronal migration in mice. *Neuron* **64**, 173–187 (2009).

195. Villar-Cerviño, V. *et al.* Contact repulsion controls the dispersion and final distribution of Cajal-Retzius cells. *Neuron* **77**, 457–471 (2013).
196. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
197. Kwan, K. Y., Šestan, N. & Anton, E. S. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development* **139**, 1535–1546 (2012).
198. Evsyukova, I., Plestant, C. & Anton, E. S. Integrative mechanisms of oriented neuronal migration in the developing brain. *Annu. Rev. Cell Dev. Biol.* **29**, 299–353 (2013).
199. Lodato, S. & Arlotta, P. Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
200. Srinivasan, K. *et al.* A network of genetic repression and derepression specifies projection fates in the developing neocortex. *Proc. Natl. Acad. Sci. USA* **109**, 19071–19078 (2012).
201. Britanova, O. *et al.* *Satb2* is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron* **57**, 378–392 (2008).
202. Dominguez, M. H., Ayoub, A. E. & Rakic, P. POU-III Transcription Factors (*Brn1*, *Brn2*, and *Oct6*) Influence Neurogenesis, Molecular Identity, and Migratory Destination of Upper-Layer Cells of the Cerebral Cortex. *Cereb. Cortex* **23**, 2632–2643 (2012).
203. Maricic, T. *et al.* A recent evolutionary change affects a regulatory element in the human *FOXP2* gene. *Mol. Biol. Evol.* **30**, 844–852 (2013).
204. Clovis, Y. M., Enard, W., Marinaro, F., Huttner, W. B. & De Pietri Tonelli, D. Convergent repression of *Foxp2* 3'UTR by miR-9 and miR-132 in embryonic mouse neocortex: Implications for radial migration of neurons. *Development* **139**, 3332–3342 (2012).
205. Garcia-Calero, E., Botella-Lopez, A., Bahamonde, O., Perez-Balaguer, A. & Martinez, S. *FoxP2* protein levels regulate cell morphology changes and migration patterns in the vertebrate developing telencephalon. *Brain Struct. Funct.* **221**, 2905–2917 (2016).
206. Teramitsu, I., Kudo, L. C., London, S. E., Geschwind, D. H. & White, S. A. Parallel *FoxP1* and *FoxP2* expression in songbird and human brain predicts functional interaction. *J. Neurosci.* **24**, 3152–3163 (2004).
207. Hamdan, F. F. *et al.* De novo mutations in *FOXP1* in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* **87**, 671–678 (2010).
208. Miyoshi, G. & Fishell, G. Dynamic *FoxG1* expression coordinates the integration of multipolar pyramidal neuron precursors into the cortical plate. *Neuron* **74**, 1045–1058 (2012).
209. Baek, S. T. *et al.* An AKT3-*FOXG1*-reelin network underlies defective migration in human focal malformations of cortical development. *Nat. Med.* **21**, 1445–1454 (2015).
210. Bayam, E. *et al.* Genome-wide target analysis of *NEUROD2* provides new insights into regulation of cortical projection neuron migration and differentiation. *BMC*



- Genomics* **16**, 681 (2015).
211. Narayanan, R. *et al.* Loss of BAF (mSWI/SNF) complexes causes global transcriptional and chromatin state changes in forebrain development. *Cell Rep.* **13**, 1842–1854 (2015).
  212. Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).
  213. Wiegreffe, C. *et al.* Bcl11a (Ctip1) controls migration of cortical projection neurons through regulation of Sema3c. *Neuron* **87**, 311–325 (2015).
  214. Rajcan-Separovic, E. *et al.* Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *J. Med. Genet.* **44**, 269–276 (2007).
  215. Dias, C. *et al.* BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. *Am. J. Hum. Genet.* **99**, 253–274 (2016).
  216. Marín, O., Valiente, M., Ge, X. & Tsai, L.-H. Guiding neuronal cell migrations. *Cold Spring Harb. Perspect. Biol.* **2**, a001834 (2010).
  217. Cheng, P. & Poo, M. Early events in axon/dendrite polarization. *Annu. Rev. Neurosci.* **35**, 181–201 (2012).
  218. Barnes, A. P. & Polleux, F. Establishment of axon-dendrite polarity in developing neurons. *Annu. Rev. Neurosci.* **32**, 347–381 (2009).
  219. Lefebvre, J. L., Sanes, J. R. & Kay, J. N. Development of dendritic form and function. *Annu. Rev. Cell Dev. Biol.* **31**, 741–777 (2015).
  220. de Wit, J., Hong, W., Luo, L. & Ghosh, A. Role of leucine-rich repeat proteins in the development and function of neural circuits. *Annu. Rev. Cell Dev. Biol.* **27**, 697–729 (2011).
  221. de Wit, J. & Ghosh, A. Specification of synaptic connectivity by cell surface interactions. *Nat. Rev. Neurosci.* **17**, 22–35 (2016).
  222. Puram, S. V & Bonni, A. Cell-intrinsic drivers of dendrite morphogenesis. *Development* **140**, 4657–4671 (2013).
  223. Hoshiba, Y. *et al.* Sox11 balances dendritic morphogenesis with neuronal migration in the developing cerebral cortex. *J. Neurosci.* **36**, 5775–5784 (2016).
  224. Chen, L. Y., Jiang, M., Zhang, B., Gokce, O. & Südhof, T. C. Conditional deletion of all neurexins defines diversity of essential synaptic organizer functions for neurexins. *Neuron* **94**, 611–625 (2017).
  225. Kwon, H.-B. & Sabatini, B. L. Glutamate induces de novo growth of functional spines in developing cortex. *Nature* **474**, 100–104 (2011).
  226. Oh, W. C., Lutz, S., Castillo, P. E. & Kwon, H.-B. De novo synaptogenesis induced by GABA in the developing mouse cortex. *Science* **353**, 1037–1040 (2016).
  227. Kizuka, Y. & Oka, S. Regulated expression and neural functions of human natural killer-1 (HNK-1) carbohydrate. *Cell. Mol. Life Sci.* **69**, 4135–4147 (2012).
  228. Blaustein, M. P. & Lederer, W. J. Sodium/calcium exchange: Its physiological implications. *Physiol. Rev.* **79**, 763–854 (1999).
  229. Altimimi, H. F. & Schnetkamp, P. P. M. Na<sup>+</sup>/Ca<sup>2+</sup>-K<sup>+</sup> exchangers (NCKX): Functional properties and physiological roles. *Channels* **1**, 62–69 (2014).
  230. Napolioni, V., Persico, A. M., Porcelli, V. & Palmieri, L. The mitochondrial

- aspartate/glutamate carrier AGC1 and calcium homeostasis: Physiological links and abnormalities in autism. *Mol. Neurobiol.* **44**, 83–92 (2011).
231. Rivera, C. *et al.* The K<sup>+</sup>/Cl<sup>-</sup> co-transporter KCC2 renders GABA hyperpolarizing during neuronal maturation. *Nature* **397**, 251–255 (1999).
  232. Diering, G. H., Mills, F., Bamji, S. X. & Numata, M. Regulation of dendritic spine growth through activity-dependent recruitment of the brain-enriched Na<sup>+</sup>/H<sup>+</sup> exchanger NHE5. *Mol. Biol. Cell* **22**, 2246–2257 (2011).
  233. Sakai, R., Cohen, D. M., Henry, J. F., Burrin, D. G. & Reeds, P. J. Leucine-nitrogen metabolism in the brain of conscious rats: Its role as a nitrogen carrier in glutamate synthesis in glial and neuronal metabolic compartments. *J. Neurochem.* **88**, 612–622 (2004).
  234. Zaia, K. A. & Reimer, R. J. Synaptic vesicle protein NTT4/XT1 (SLC6A17) catalyzes Na<sup>+</sup>-coupled neutral amino acid transport. *J. Biol. Chem.* **284**, 8439–8448 (2009).
  235. Furberg, H. *et al.* Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
  236. Srour, M. *et al.* Dysfunction of the cerebral glucose transporter SLC45A1 in individuals with intellectual disability and epilepsy. *Am. J. Hum. Genet.* **100**, 824–830 (2017).
  237. Sanders, A. F. *Elements of Human Performance: Reaction Processes and Attention in Human Skill.* (Erlbaum, 1998).
  238. Edwards, R. H. The neurotransmitter cycle and quantal size. *Neuron* **55**, 835–858 (2007).
  239. Wilhelm, B. G. *et al.* Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* **344**, 1023–1028 (2014).
  240. Südhof, T. C. The presynaptic active zone. *Neuron* **75**, 11–25 (2012).
  241. Hutagalung, A. H. & Novick, P. J. Role of Rab GTPases in membrane traffic and cell physiology. *Physiol. Rev.* **91**, 119–149 (2011).
  242. von Mollard, G., Stahl, B., Khokhlatchev, A., Südhof, T. C. & Jahn, R. Rab3C is a synaptic vesicle protein that dissociates from synaptic vesicles after stimulation of exocytosis. *J. Biol. Chem.* **269**, 10971–10974 (1994).
  243. Schlüter, O. M., Schmitz, F., Jahn, R., Rosenmund, C. & Südhof, T. C. A complete genetic analysis of neuronal Rab3 function. *J. Neurosci.* **24**, 6629–6637 (2004).
  244. Sakane, A. *et al.* Rab3 GTPase-activating protein regulates synaptic transmission and plasticity through the inactivation of Rab3. *Proc. Natl. Acad. Sci. USA* **103**, 10029–10034 (2006).
  245. Dolphin, A. C. Calcium channel auxiliary  $\alpha_2\delta$  and  $\beta$  subunits: Trafficking and one step beyond. *Nat. Rev. Neurosci.* **13**, 542–555 (2012).
  246. Buraei, Z. & Yang, J. The  $\beta$  subunit of voltage-gated Ca<sup>2+</sup> channels. *Physiol. Rev.* **90**, 1461–1506 (2010).
  247. Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
  248. Scales, S. J., Hesser, B. A., Masuda, E. S. & Scheller, R. H. Amisyn, a novel syntaxin-binding protein that may regulate SNARE complex assembly. *J. Biol. Chem.* **277**, 28271–28279 (2002).

249. Constable, J. R. L., Graham, M. E., Morgan, A. & Burgoyne, R. D. Amisyn regulates exocytosis and fusion pore stability by both syntaxin-dependent and syntaxin-independent mechanisms. *J. Biol. Chem.* **280**, 31615–31623 (2005).
250. Verhage, M. *et al.* Synaptic assembly of the brain in the absence of neurotransmitter secretion. *Science* **287**, 864–869 (2000).
251. Lee, H.-K. *et al.* Dynamic Ca<sup>2+</sup>-dependent stimulation of vesicle fusion by membrane-anchored synaptotagmin 1. *Science* **328**, 757–760 (2010).
252. Zhou, Q. *et al.* Architecture of the synaptotagmin-SNARE machinery for neuronal exocytosis. *Nature* **525**, 62–67 (2015).
253. Kavalali, E. T. The mechanisms and functions of spontaneous neurotransmitter release. *Nat. Rev. Neurosci.* **16**, 5–16 (2015).
254. Bacaj, T. *et al.* Synaptotagmin-1 and synaptotagmin-7 trigger synchronous and asynchronous phases of neurotransmitter release. *Neuron* **80**, 947–959 (2013).
255. Jackman, S. L., Turecek, J., Belinsky, J. E. & Regehr, W. G. The calcium sensor synaptotagmin 7 is required for synaptic facilitation. *Nature* **529**, 88–91 (2016).
256. Ryu, J.-K. *et al.* Spring-loaded unraveling of a single SNARE complex by NSF in one round of ATP turnover. *Science* **347**, 1485–1489 (2015).
257. Zhao, M. *et al.* Mechanistic insights into the recycling machine of the SNARE complex. *Nature* **518**, 61–67 (2015).
258. McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.* **12**, 517–533 (2011).
259. Kelly, B. T. *et al.* AP2 controls clathrin polymerization with a membrane-activated switch. *Science* **345**, 459–463 (2014).
260. Henley, J. M. & Wilkinson, K. A. Synaptic AMPA receptor composition in development, plasticity and disease. *Nat. Rev. Neurosci.* **17**, 337–350 (2016).
261. Herguedas, B. *et al.* Structure and organization of heteromeric AMPA-type glutamate receptors. *Science* **352**, aad3873 (2016).
262. Jackson, A. C. & Nicoll, R. A. The expanding social network of ionotropic glutamate receptors: TARPs and other transmembrane auxiliary subunits. *Neuron* **70**, 178–199 (2011).
263. Ben-Yaacov, A. *et al.* Molecular mechanism of AMPA receptor modulation by TARP/stargazin. *Neuron* **93**, 1126–1137 (2017).
264. Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–96 (2010).
265. Ascoli, G. A. *Trees of the Brain, Roots of the Mind*. (MIT Press, 2015).
266. Miller, P. S. & Aricescu, A. R. Crystal structure of a human GABA<sub>A</sub> receptor. *Nature* **512**, 270–275 (2014).
267. Olsen, R. W. & Sieghart, W. GABA<sub>A</sub> receptors: Subtypes provide diversity of function and pharmacology. *Neuropharmacology* **56**, 141–148 (2009).
268. Tyagarajan, S. K. & Fritschy, J.-M. Gephyrin: A master regulator of neuronal function? *Nat. Rev. Neurosci.* **15**, 141–156 (2014).
269. Um, J. W. *et al.* IQ motif and SEC7 domain-containing protein 3 (IQSEC3) interacts with gephyrin to promote inhibitory synapse formation. *J. Biol. Chem.* **291**, 10119–10130 (2016).
270. Hassaine, G. *et al.* X-ray structure of the mouse serotonin 5-HT<sub>3</sub> receptor. *Nature* **512**, 276–281 (2014).

271. Lummis, S. C. R. 5-HT<sub>3</sub> receptors. *J. Biol. Chem.* **287**, 40239–40245 (2012).
272. Lesch, K.-P. & Waider, J. Serotonin in the modulation of neural plasticity and networks: Implications for neurodevelopmental disorders. *Neuron* **76**, 175–191 (2012).
273. Shen, H. *et al.* Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science* **355**, eaal4326 (2017).
274. Rasband, M. N. The axon initial segment and the maintenance of neuronal polarity. *Nat. Rev. Neurosci.* **11**, 552–562 (2010).
275. Kole, M. H. P. & Stuart, G. J. Signal processing in the axon initial segment. *Neuron* **73**, 235–247 (2012).
276. Lorincz, A. & Nusser, Z. Molecular identity of dendritic voltage-gated sodium channels. *Science* **328**, 906–909 (2010).
277. Fleidervish, I. A., Lasser-Ross, N., Gutnick, M. J. & Ross, W. N. Na<sup>+</sup> imaging reveals little difference in action potential-evoked Na<sup>+</sup> influx between axon and soma. *Nat. Neurosci.* **13**, 852–860 (2010).
278. Trimmer, J. S. & Rhodes, K. J. Localization of voltage-gated ion channels in mammalian brain. *Annu. Rev. Physiol.* **66**, 477–519 (2004).
279. Gutman, G. A. *et al.* International Union of Pharmacology. LIII. Nomenclature and molecular relationships of voltage-gated potassium channels. *Pharmacol. Rev.* **57**, 473–508 (2005).
280. Lai, H. C. & Jan, L. Y. The distribution and targeting of neuronal voltage-gated ion channels. *Nat. Rev. Neurosci.* **7**, 548–562 (2006).
281. Vacher, H., Mohapatra, D. P. & Trimmer, J. S. Localization and targeting of voltage-dependent ion channels in mammalian central neurons. *Physiol. Rev.* **88**, 1407–1447 (2008).
282. Nusser, Z. in *Dendrites* (eds. Stuart, G., Spruston, N. & Häusser, M.) 191–216 (Oxford University Press, 2016).
283. Li, Y., Um, S. Y. & McDonald, T. V. Voltage-gated potassium channels: Regulation by accessory subunits. *Neuroscientist* **12**, 199–210 (2006).
284. Pongs, O. & Schwarz, J. R. Ancillary subunits associated with voltage-dependent K<sup>+</sup> channels. *Physiol. Rev.* **90**, 755–796 (2010).
285. Norris, A. J., Foeger, N. C. & Nerbonne, J. M. Interdependent roles for accessory KChIP2, KChIP3, and KChIP4 subunits in the generation of Kv4-encoded IA channels in cortical pyramidal neurons. *J. Neurosci.* **30**, 13644–13655 (2010).
286. Lee, U. S. & Cui, J. BK channel activation: Structural and functional insights. *Trends Neurosci.* **33**, 415–423 (2010).
287. Tao, X., Hite, R. K. & MacKinnon, R. Cryo-EM structure of the open high-conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel. *Nature* **541**, 46–51 (2017).
288. Hite, R. K., Tao, X. & MacKinnon, R. Structural basis for gating the high-conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channel. *Nature* **541**, 52–57 (2017).
289. Adelman, J. P., Maylie, J. & Sah, P. Small-conductance Ca<sup>2+</sup>-activated K<sup>+</sup> channels: Form and function. *Annu. Rev. Physiol.* **74**, 245–269 (2012).
290. Lorincz, A., Notomi, T., Tamas, G., Shigemoto, R. & Nusser, Z. Polarized and compartment-dependent distribution of HCN1 in pyramidal cell dendrites. *Nat. Neurosci.* **5**, 1185–1193 (2002).
291. Wettschureck, N. & Offermanns, S. Mammalian G proteins and their cell type

- specific functions. *Physiol. Rev.* **85**, 1159–1204 (2005).
292. Kimple, A. J., Bosch, D. E., Giguère, P. M. & Siderovski, D. P. Regulators of G-protein signaling and their Gα substrates: Promises and challenges in their use as drug discovery targets. *Pharmacol. Rev.* **63**, 728–749 (2011).
  293. Jayaraman, M., Zhou, H., Jia, L., Cain, M. D. & Blumer, K. J. R9AP and R7BP: Traffic cops for the RGS7 family in phototransduction and neuronal GPCR signaling. *Trends Pharmacol. Sci.* **30**, 17–24 (2009).
  294. Thal, D. M. *et al.* Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. *Nature* **531**, 335–340 (2016).
  295. Hibino, H. *et al.* Inwardly rectifying potassium channels: Their structure, function, and physiological roles. *Physiol. Rev.* **90**, 291–366 (2010).
  296. Pidoux, G. & Taskén, K. Specificity and spatial dynamics of protein kinase A signaling organized by A-kinase-anchoring proteins. *J. Mol. Endocrinol.* **44**, 271–284 (2010).
  297. Conti, M. & Beavo, J. Biochemistry and physiology of cyclic nucleotide phosphodiesterases: Essential components in cyclic nucleotide signaling. *Annu. Rev. Biochem.* **76**, 481–511 (2007).
  298. Winder, D. G. & Sweatt, J. D. Roles of serine/threonine phosphatases in hippocampal synaptic plasticity. *Nat. Rev. Neurosci.* **2**, 461–474 (2001).
  299. Hussain, N. K. & Hugarir, R. L. in *Dendrites* (eds. Stuart, G., Spruston, N. & Häusser, M.) 159–189 (Oxford University Press, 2016).
  300. Tang, A.-H. *et al.* A trans-synaptic nanocolumn aligns neurotransmitter release to receptors. *Nature* **536**, 210–214 (2016).
  301. Zhu, J., Shang, Y. & Zhang, M. Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. *Nat. Rev. Neurosci.* **17**, 209–223 (2016).
  302. Deng, F., Price, M. G., Davis, C. F., Mori, M. & Burgess, D. L. Stargazin and other transmembrane AMPA receptor regulating proteins interact with synaptic scaffolding protein MAGI-2 in brain. *J. Neurosci.* **26**, 7875–7884 (2006).
  303. Danielson, E. *et al.* S-SCAM/MAGI-2 is an essential synaptic scaffolding molecule for the GluA2-containing maintenance pool of AMPA receptors. *J. Neurosci.* **32**, 6967–6980 (2012).
  304. Monteiro, P. & Feng, G. SHANK proteins: Roles at the synapse and in autism spectrum disorder. *Nat. Rev. Neurosci.* **18**, 147–157 (2017).
  305. Leblond, C. S. *et al.* Meta-analysis of SHANK mutations in autism spectrum disorders: A gradient of severity in cognitive impairments. *PLoS Genet.* **10**, e1004580 (2014).
  306. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
  307. Sanders, S. J. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
  308. Bidinosti, M. *et al.* CLK2 inhibition ameliorates autistic features associated with SHANK3 deficiency. *Science* **351**, 1199–1203 (2016).
  309. Sanderson, J. L. & Dell’Acqua, M. L. AKAP signaling complexes in regulation of excitatory synaptic plasticity. *Neuroscientist* **17**, 321–336 (2011).
  310. Rotty, J. D., Wu, C. & Bear, J. E. New insights into the regulation and cellular

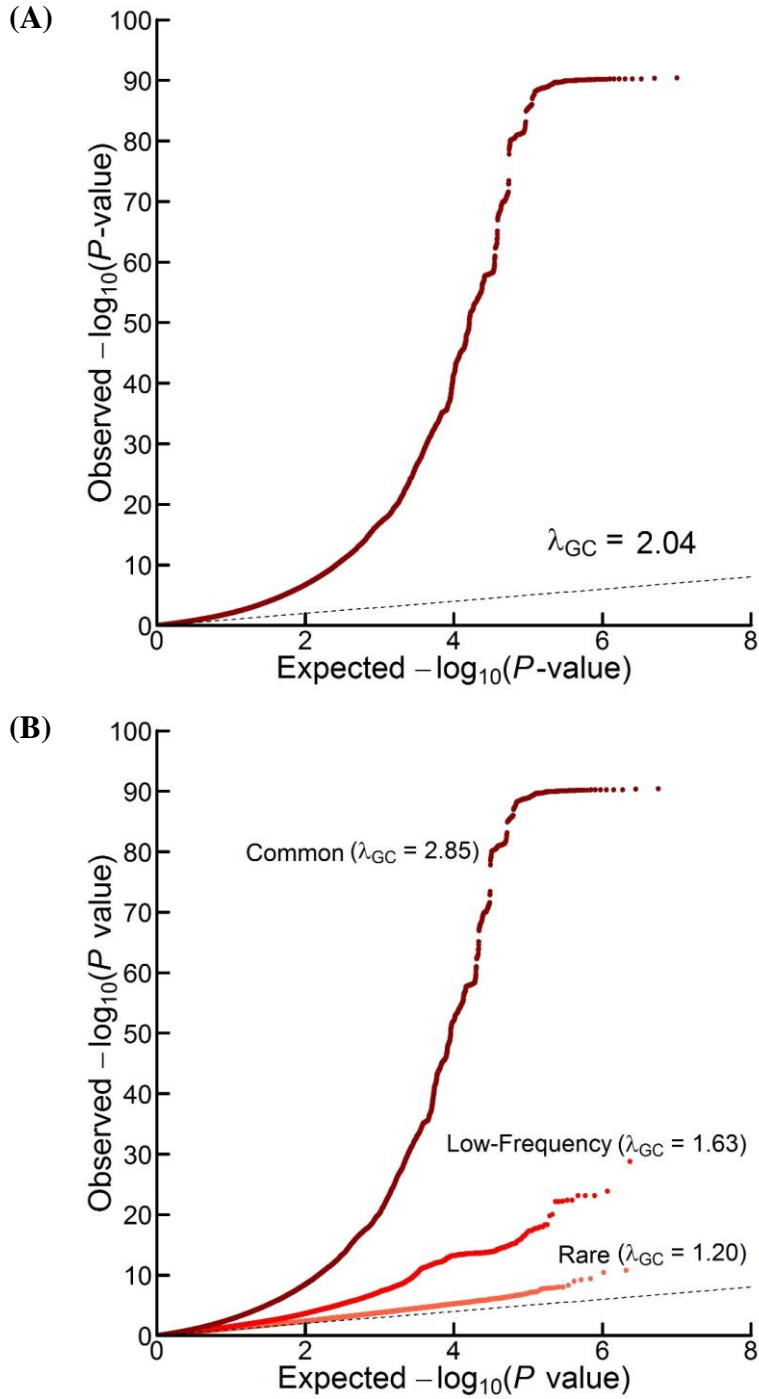
- functions of the ARP2/3 complex. *Nat. Rev. Mol. Cell Biol.* **14**, 7–12 (2013).
311. Kim, I. H. *et al.* Disruption of Arp2/3 results in asymmetric structural plasticity of dendritic spines and progressive synaptic and behavioral abnormalities. *J. Neurosci.* **33**, 6081–6092 (2013).
  312. Herring, B. E. & Nicoll, R. A. Kalirin and Trio proteins serve critical roles in excitatory synaptic transmission and LTP. *Proc. Natl. Acad. Sci. USA* **113**, 2264–2269 (2016).
  313. Yan, Y., Eipper, B. A. & Mains, R. E. Kalirin-9 and Kalirin-12 play essential roles in dendritic outgrowth and branching. *Cereb. Cortex* **25**, 3487–3501 (2015).
  314. Murakoshi, H., Wang, H. & Yasuda, R. Local, persistent activation of Rho GTPases during plasticity of single dendritic spines. *Nature* **472**, 100–104 (2011).
  315. Tanaka, J.-I. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319**, 1683–1687 (2008).
  316. Govindarajan, A., Israely, I., Huang, S.-Y. & Tonegawa, S. The dendritic branch is the preferred integrative unit for protein synthesis-dependent LTP. *Neuron* **69**, 132–146 (2011).
  317. Wu, D. *et al.* Postsynaptic synaptotagmins mediate AMPA receptor exocytosis during LTP. *Nature* **544**, 316–321 (2017).
  318. Collingridge, G. L., Peineau, S., Howland, J. G. & Wang, Y. T. Long-term depression in the CNS. *Nat. Rev. Neurosci.* **11**, 459–473 (2010).
  319. Connor, S. A. & Wang, Y. T. A place at the table: LTD as a mediator of memory genesis. *Neuroscientist* **22**, 359–371 (2016).
  320. Greengard, P. The neurobiology of slow synaptic transmission. *Science* **294**, 1024–1030 (2001).
  321. Heroes, E. *et al.* The PP1 binding code: A molecular-lego strategy that governs specificity. *FEBS J.* **280**, 584–595 (2013).
  322. Niswender, C. M. & Conn, P. J. Metabotropic glutamate receptors: Physiology, pharmacology, and disease. *Annu. Rev. Pharmacol. Toxicol.* **50**, 295–322 (2010).
  323. Nicoletti, F. *et al.* Metabotropic glutamate receptors: From the workbench to the bedside. *Neuropharmacology* **60**, 1017–1041 (2011).
  324. Collingridge, G. L., Manahan-Vaughan, D., Nicoletti, F. & Schoepp, D. D. Metabotropic glutamate receptors, 5 years on. *Neuropharmacology* **115**, 1–3 (2017).
  325. Yuzaki, M. & Aricescu, A. R. A GluD coming-of-age story. *Trends Neurosci.* **40**, 138–150 (2017).
  326. Koleske, A. J. Molecular mechanisms of dendrite stability. *Nat. Rev. Neurosci.* **14**, 536–550 (2013).
  327. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* **49**, 1107–1112 (2017).
  328. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).
  329. Wang, G. *et al.* CaV3.2 calcium channels control NMDA receptor-mediated transmission: a new mechanism for absence epilepsy. *Genes Dev.* **29**, 1535–51 (2015).
  330. Yalin, Ö. Genes and molecular mechanisms involved in the epileptogenesis of

- idiopathic absence epilepsies. *Seizure* (2012). doi:10.1016/j.seizure.2011.12.002
331. Rietveld, C. A. *et al.* Common Genetic Variants Associated with Cognitive Performance Identified Using Proxy-Phenotype Method. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13790–13794 (2014).
  332. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
  333. Ardlie, K. G. *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
  334. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
  335. Vilhjálmsdóttir, B. J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
  336. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
  337. Altshuler, D. M., Gibbs, R. A. & Peltonen, L. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
  338. Harris, K. M., Halpern, C. T., Haberstick, B. C. & Smolen, A. The National Longitudinal Study of Adolescent Health (Add Health) sibling pairs data. *Twin Res. Hum. Genet.* **16**, 391–8 (2013).
  339. Sonnega, A. *et al.* Cohort profile: The Health and Retirement Study (HRS). *Int. J. Epidemiol.* **43**, 576–585 (2014).
  340. Dunn, L. M. & Dunn, D. M. Peabody Picture Vocabulary Test,. *Summ. Shute. Inst.* **30**, 1–8 (2007).
  341. Cameron, R. C. & Miller, D. L. A Practitioner’s Guide to Cluster-Robust Inference. *J. Hum. Resour.* **50**, 317–372 (2015).
  342. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
  343. Vassos, E. *et al.* An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis. *Biol. Psychiatry* **81**, 470–477 (2017).
  344. Belsky, D. W. *et al.* Development and evaluation of a genetic risk score for obesity. *Biodemography Soc. Biol.* **59**, 85–100 (2013).
  345. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
  346. Hill, W. D., Davies, G., McIntosh, A. M., Gale, C. R. & Deary, I. J. A combined analysis of genetically correlated traits identifies 107 loci associated with intelligence. *bioRxiv* (2017). doi:10.1101/160291
  347. Davies, G. *et al.* Ninety-nine independent genetic loci influencing general cognitive function include genes associated with brain health and structure (N = 280,360). *bioRxiv* (2017).
  348. Herd, P., Carr, D. & Roan, C. Cohort profile: Wisconsin longitudinal study (WLS). *Int. J. Epidemiol.* **43**, 34–41 (2014).
  349. Hauser, R. M. & Willis, R. J. Survey Design and Methodology in the Health and Retirement Study and the Wisconsin Longitudinal Study. *Population and Development Review* **30**, 209–235
  350. Aitkin, M. A. Correlation in a singly truncated bivariate normal distribution. *Psychometrika* **29**, 263–270 (1964).

351. Ofstedal, M. B. *et al.* HRS/AHEAD Documentation Report Documentation of Cognitive Functioning Measures in the Health and Retirement Study. (2005).
352. Henmon, V. A. C. Henmon-Nelson Tests of Mental Ability, High School Examination-Grades 7 to 12-Forms A, B, and C. Teacher's Manual. (1946).
353. Fain, G. L. *Molecular and Cellular Physiology of Neurons*. (Harvard University Press). doi:10.1017/CBO9781107415324.004
354. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
355. Lyall, D. M. *et al.* Cognitive Test Scores in UK Biobank: Data Reduction in 480,416 Participants and Longitudinal Stability in 20,346 Participants. *PLoS One* **11**, e0154222 (2016).
356. Tropf, F. C. *et al.* Hidden heritability due to heterogeneity across seven populations. *Nat. Hum. Behav.* (2017). doi:10.1038/s41562-017-0195-1
357. Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.* **18**, 86 (2017).
358. Robinson, M. J. & Cobb, M. H. Mitogen-activated protein kinase pathways. *Curr. Opin. Cell Biol.* **9**, 180–186 (1997).
359. Thomas, G. M. & Huganir, R. L. MAPK cascade signalling and synaptic plasticity. *Nat. Rev. Neurosci.* **5**, 173–183 (2004).
360. Hu, W. *et al.* Distinct contributions of Na<sub>v</sub>1.6 and Na<sub>v</sub>1.2 in action potential initiation and backpropagation. *Nat. Neurosci.* **12**, 996–1002 (2009).

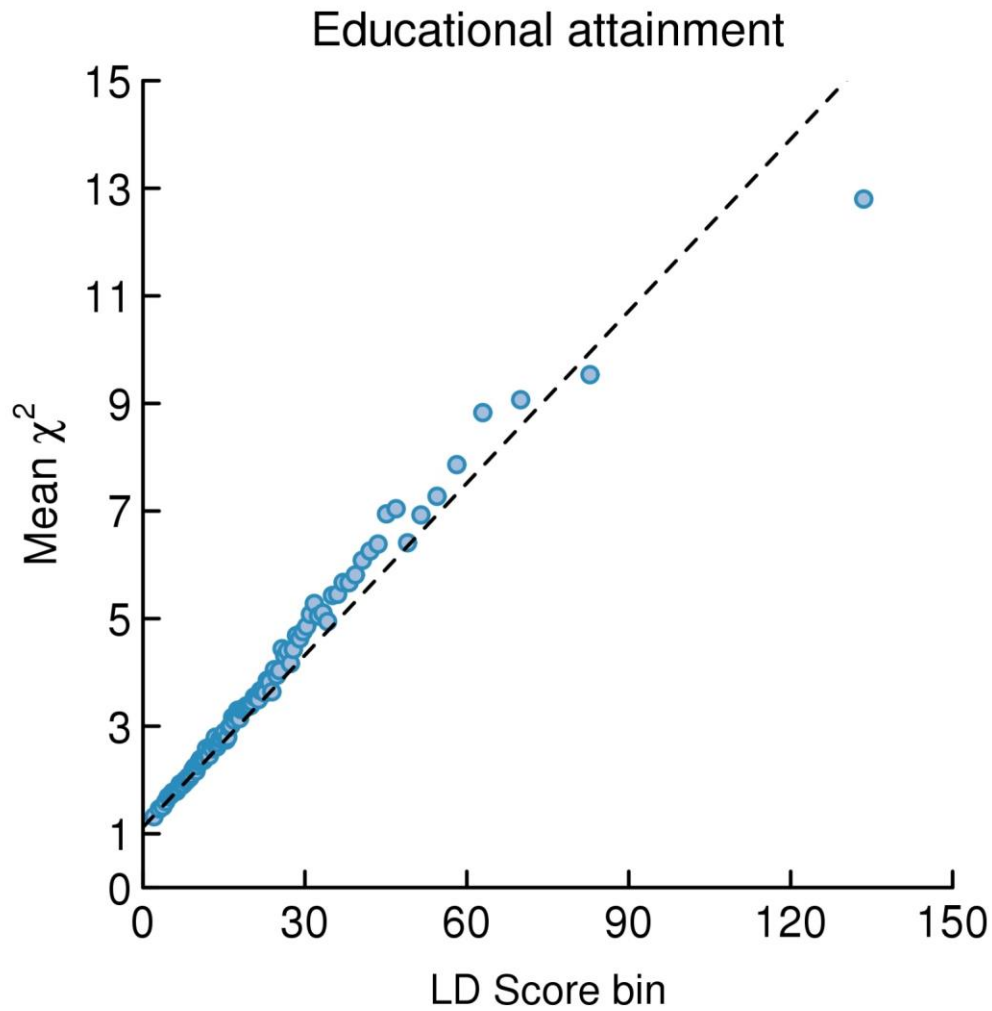


## **Supplementary Figures**



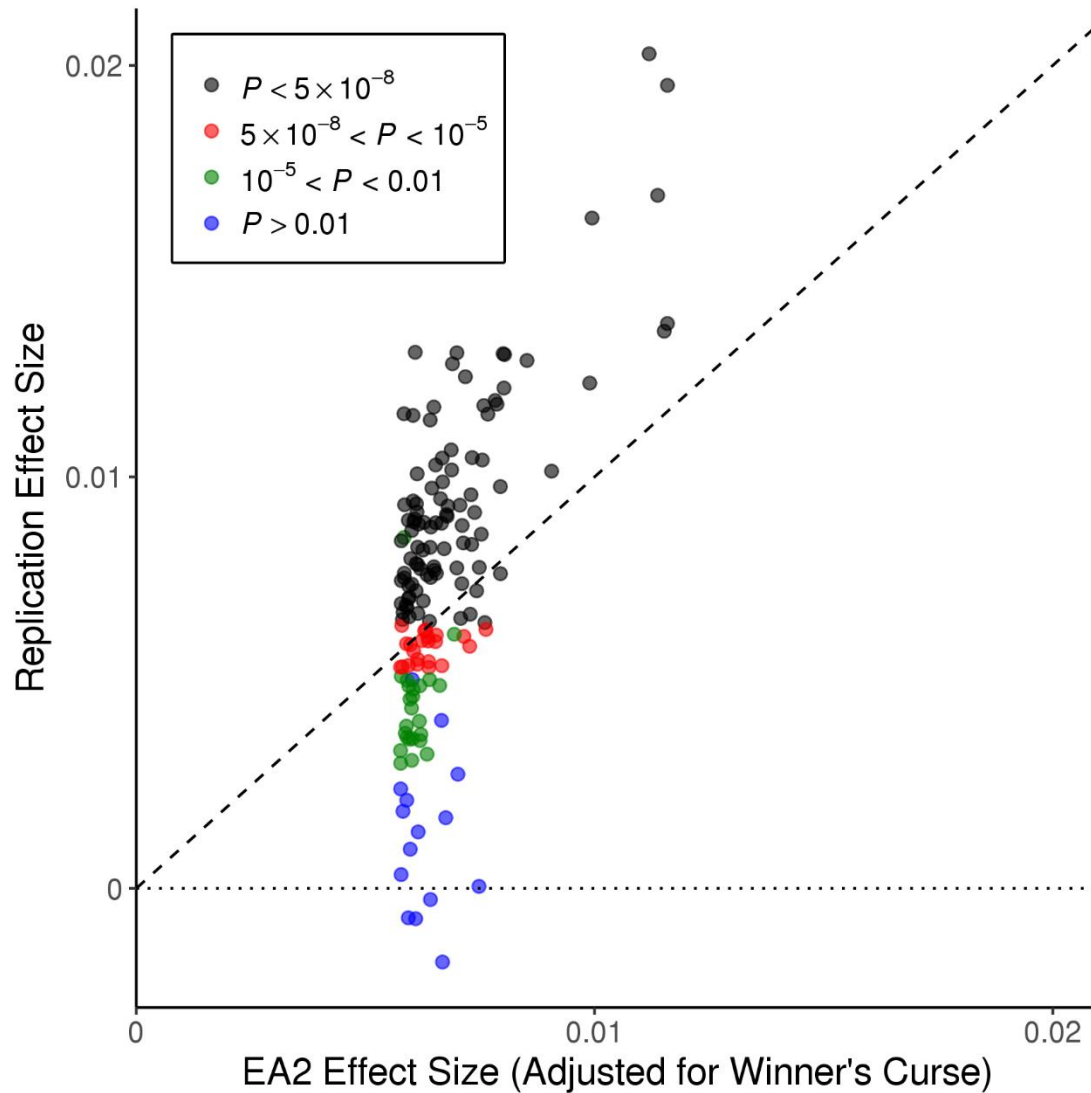
**Supplementary Figure 1. Quantile-quantile Plots from Meta-analysis of *EduYears* ( $N = 1,131,881$ ).**

(A) All SNPs. (B) SNPs grouped by minor allele frequency (MAF); rare ( $< 1\%$ ), low-frequency ( $1\text{-}5\%$ ) and common ( $> 5\%$ ).



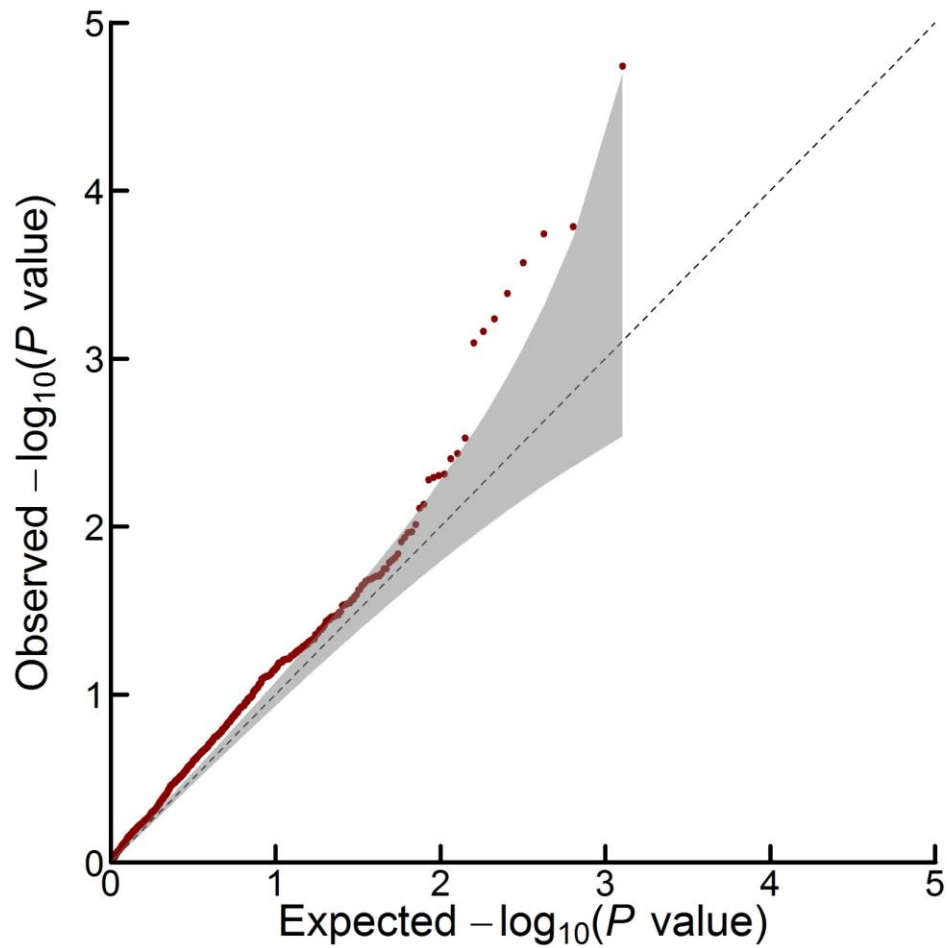
**Supplementary Figure 2. LD Score Plot from Meta-analysis of *EduYears* ( $N = 1,131,881$ ).**

Each point represents an LD Score quantile. The  $x$  and  $y$  coordinates of the point are the mean LD Score and the mean  $\chi^2$  statistic of SNPs in that quantile. The fact that the intercept is close to one and that  $\chi^2$  statistics increase linearly with the LD Scores of the SNPs suggests that the bulk of the inflation in the  $\chi^2$  statistics is due to true polygenic signal.



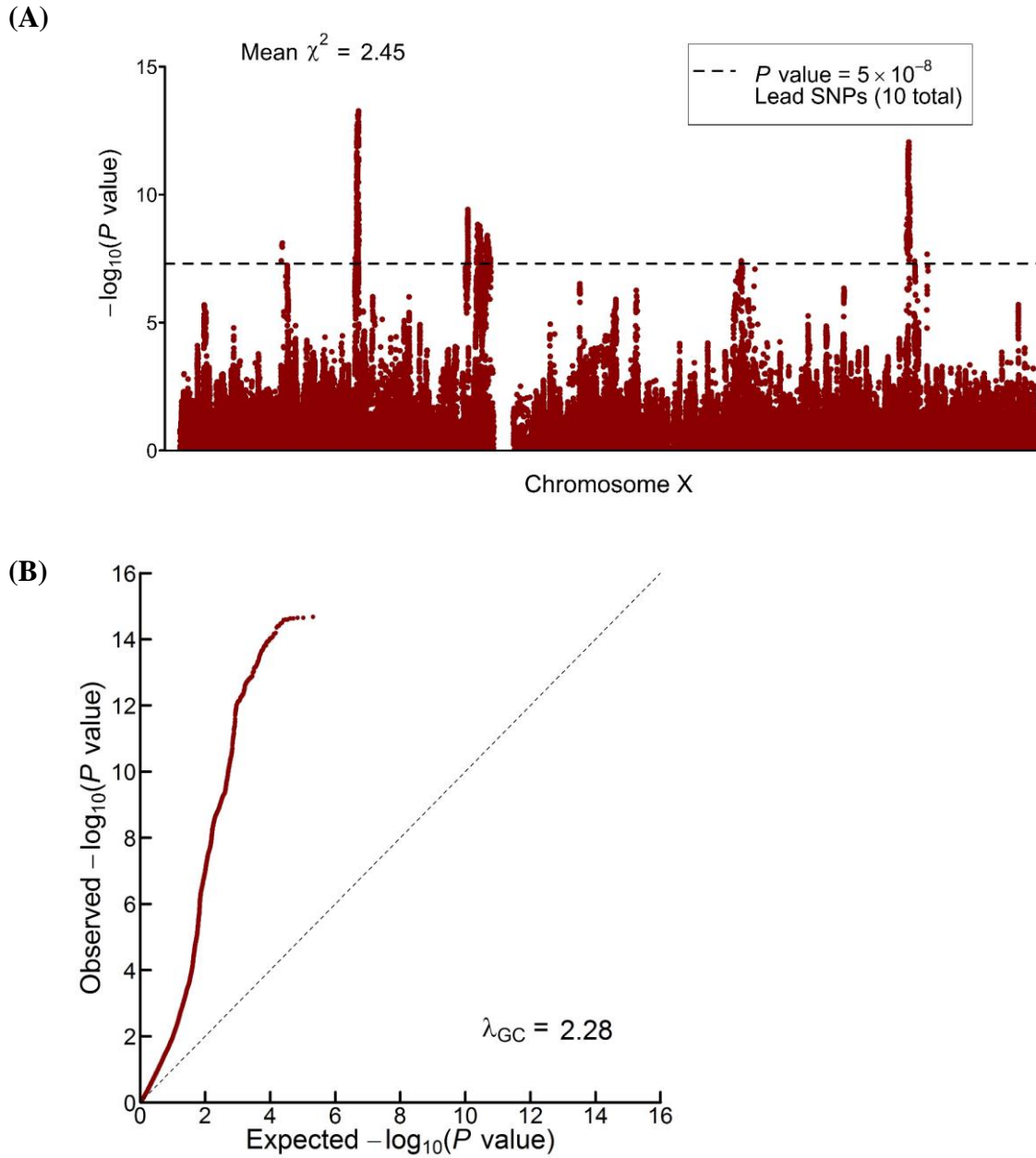
### Supplementary Figure 3. Replication of EA2 Lead SNPs

We examined the out-of-sample replicability of the lead SNPs identified at genome-wide significance in the combined-stage sample ( $N = 405,073$ ) of a published GWAS meta-analysis of *EduYears*<sup>1</sup>. The  $x$ -axis is the winner's-curse-adjusted estimate of the SNP's effect size in the previous study (calculated using shrinkage parameters estimated using summary statistics from the previous study). The  $y$ -axis is the SNP's effect size estimated from the subsample of our data ( $N = 726,808$ ) that did not contribute to the previous study's meta-analysis. All effect sizes are from a regression where phenotype and genotype have been standardized to have unit variance. The reference allele is chosen to be the allele estimated to increase EA in the previous study; therefore, all points above the dotted line have matching signs in the replication sample.



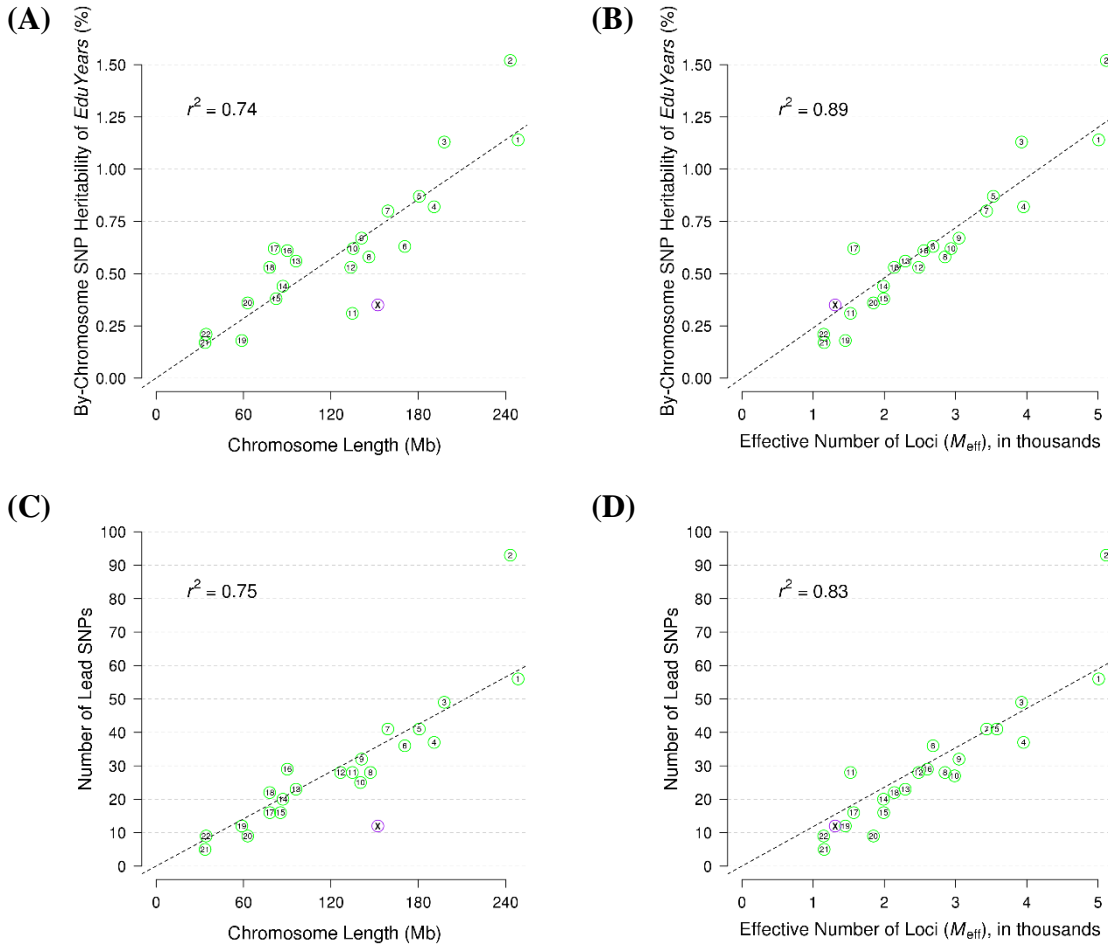
**Supplementary Figure 4. Testing for Heterogeneous Effects of Lead SNPs.**

For each SNP included in the meta-analysis, we used Cochran's Q test of heterogeneity<sup>9</sup> to test for effect-size heterogeneity across cohorts. The figure above shows the quantiles of the observed distribution of heterogeneity  $P$  values for our 1,271 lead SNPs against the distribution under the null hypothesis of homogeneous SNP effects across cohorts. The gray shaded areas in the Q–Q plots represent the 95% confidence intervals under the null hypothesis. See **Supplementary Table 2** for underlying data.



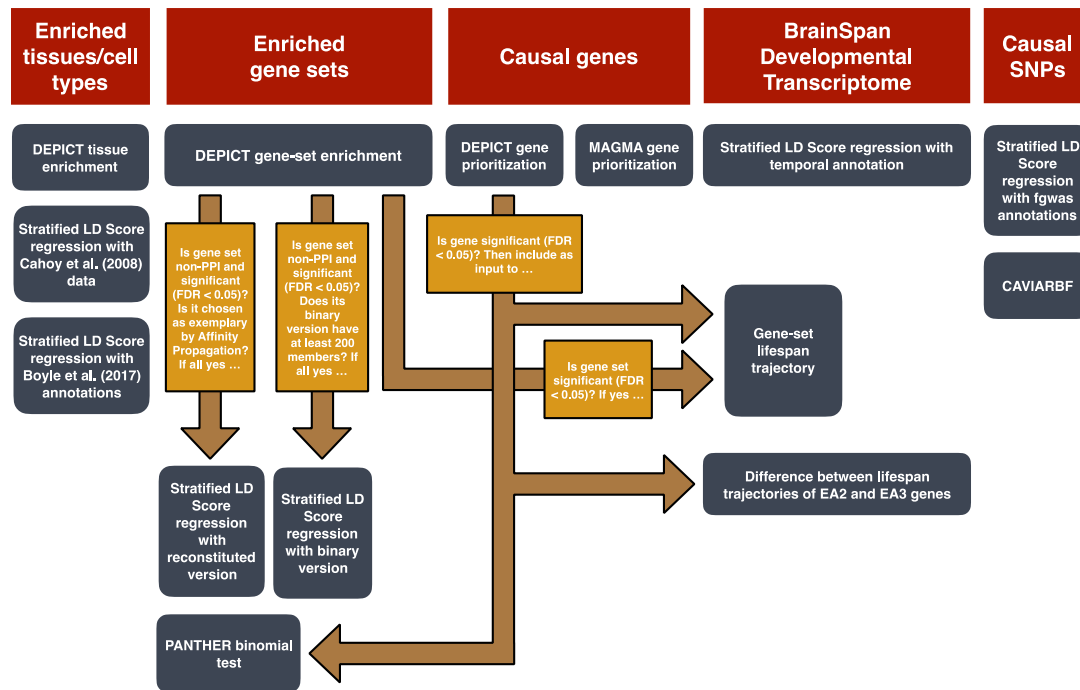
**Supplementary Figure 5. Meta-Analysis of X-Chromosomal SNPs ( $N = 694,894$ ).**

The meta-analysis was conducted by combining summary statistics from (sex-pooled) association analyses conducted in UK Biobank ( $N = 329,538$ ) and 23andMe ( $N = 365,356$ ); see **Supplementary Section 4.6** for additional details. In (A), the  $P$  values plotted are based on summary statistics adjusted for inflation using the same LD Score intercept used in the autosomal analyses. The solid line indicates the threshold for genome-wide significance ( $P = 5 \times 10^{-8}$ ). The  $P$  values in (B) are based on  $P$  values from unadjusted test statistics.



### Supplementary Figure 6. Comparison of Autosomal and X-Chromosomal Association Results.

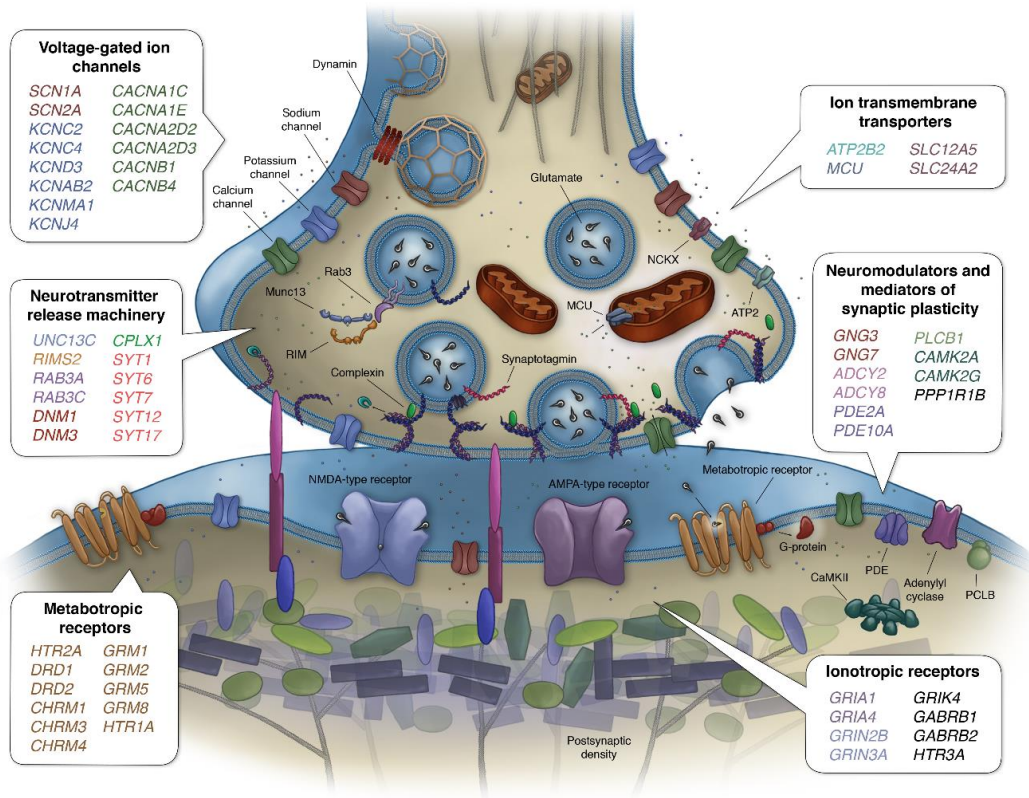
“Chromosome length” is calculated as the difference between the minimum and maximum base-pair position SNPs on the chromosome. “Effective number of loci ( $M_{\text{eff}}$ )” is calculated from the  $M$  SNPs using data from the UK Biobank. For each chromosome, “by-chromosome SNP heritability” is calculated as  $h^2 = \frac{(\overline{\chi^2} - 1)M_{\text{eff}}}{\bar{N}}$ , where  $\overline{\chi^2}$  is the mean  $\chi^2$  test statistic for that chromosome and  $\bar{N}$  is the average GWAS sample size. “Number of lead SNPs” is calculated by applying our clumping algorithm (see **Supplementary Section 1.8**) to the set of genome-wide significant SNPs; to account for the larger GWAS sample size available for the autosomes relative to the X chromosome, we inflate the standard errors for the autosomal GWAS results by  $\sqrt{\frac{808,000}{694,894}} \approx 1.079$ . The dashed line is the best fit from a regression of the points in the plot with the intercept constrained to zero. The value  $r^2$  is the squared correlation coefficient of the points in each plot. **(A)** Chromosome length vs. heritability, **(B)** Effective number of loci vs. heritability, **(C)** Chromosome length vs. number of lead SNPs, **(D)** Effective number of loci vs. number of lead SNPs.



### Supplementary Figure 7. Flowchart of Biological Annotation.

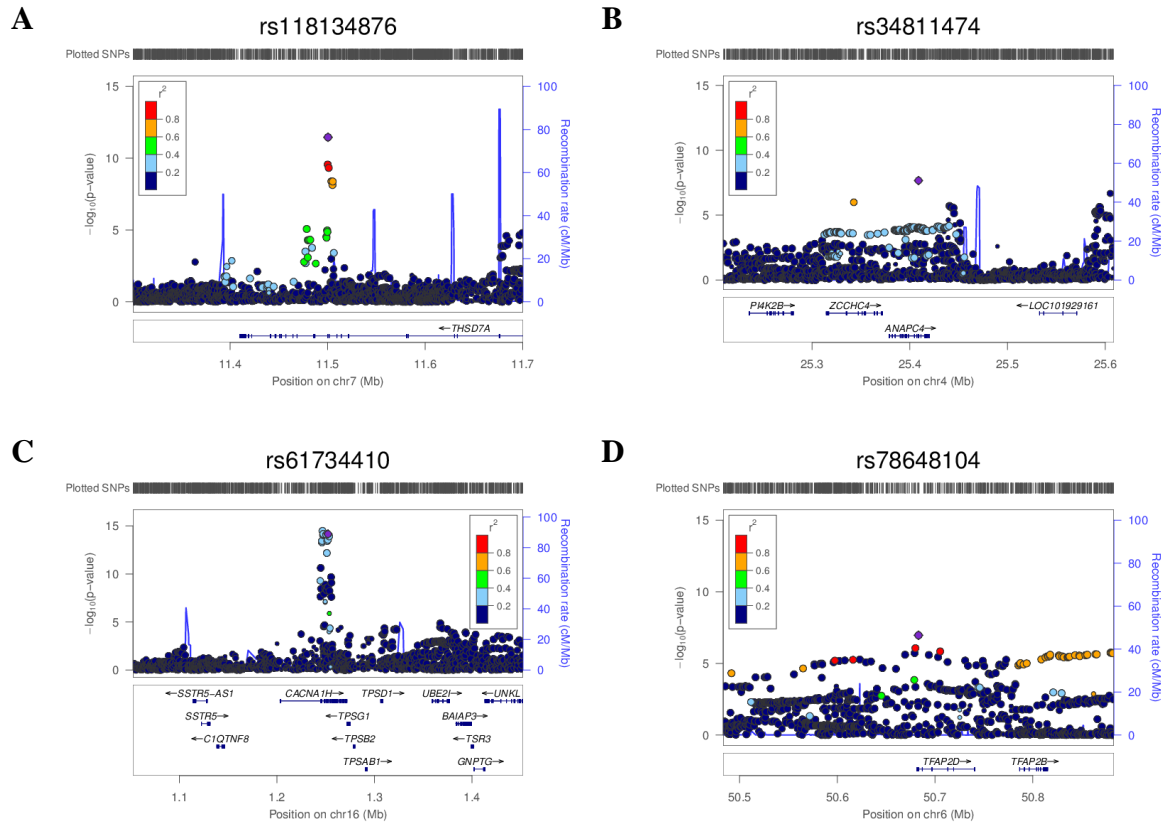
Our various biological analyses are grouped into five broad classes: (1) identification of enriched tissues/cell types, (2) identification of enriched gene sets, (3) prioritization of likely causal genes, (4) use of the BrainSpan Developmental Transcriptome to analyze when the causal genes are expressed, and (5) prioritization of likely causal SNPs. These broad classes are represented by the red boxes. Each gray box corresponds to a particular analysis and is placed under the broad class of analyses (red box) to which it belongs. Some analyses use the output of other analyses as input; such a relationship is represented by a brown arrow. The label of an arrow describes how the output of the prior analysis was filtered to produce input for later analyses. EA2, prioritized by DEPICT in Okbay et al.<sup>1</sup> and in the current study; EA3, prioritized by DEPICT in the current study but not in Okbay et al.<sup>1</sup>.





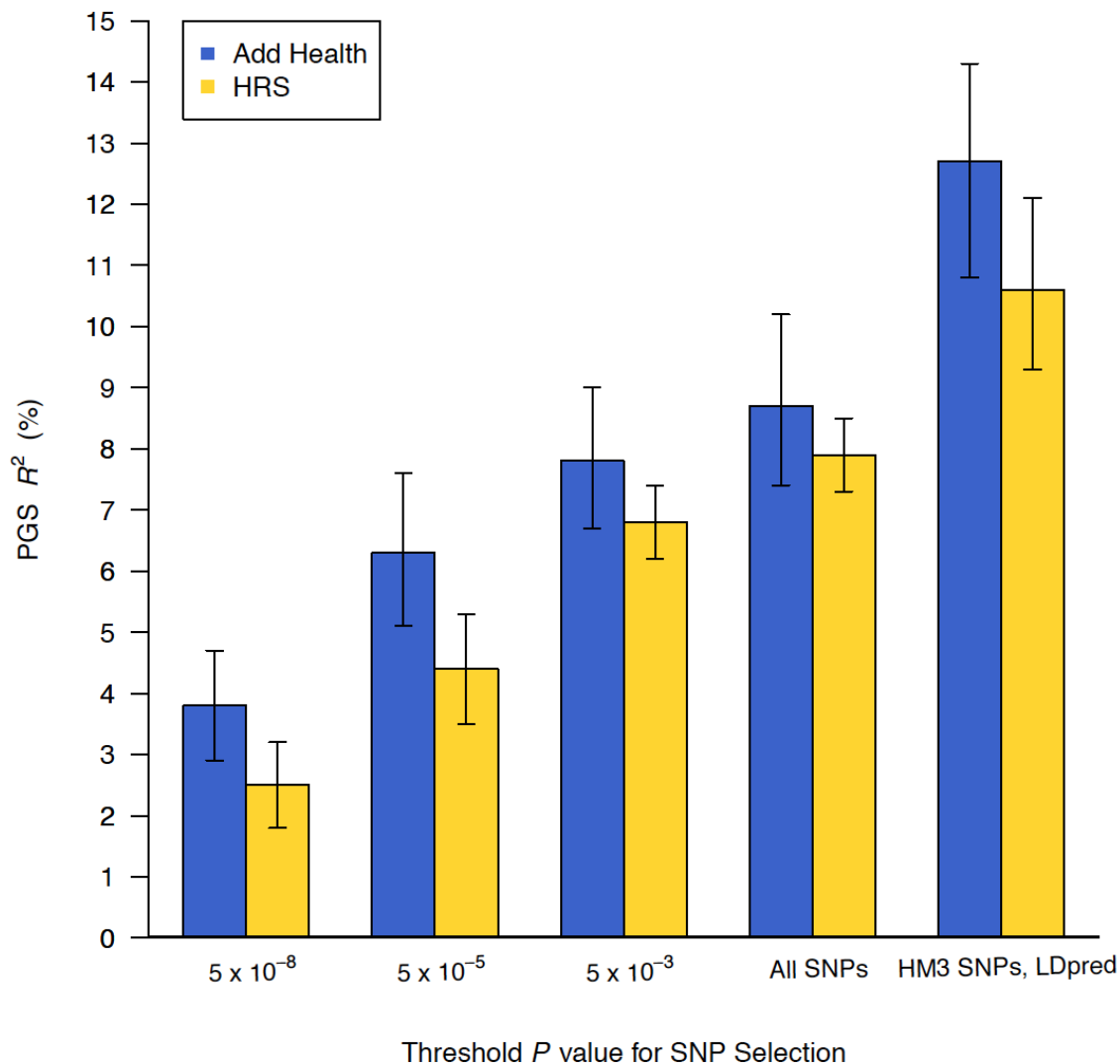
### Supplementary Figure 8. Roles of Selected Newly Prioritized Genes in Neuronal Communication.

The 59 genes listed in the figure were selected as follows. We began with the 30 gene-set clusters in **Supplementary Figure 22** and dropped those that include gene sets that were implicated in a previous study of *EduYears* (Supplementary Table 4.5.1 of Okbay et al.<sup>1</sup>). Of the 8 clusters that remained, we retained the 4 related to neuronal communication (**DAG and IP<sub>3</sub> signaling, associative learning, post NMDA receptor activation events, regulation of neurotransmitter levels**). We identified the 460 DEPICT-prioritized genes belonging to the exemplary gene sets representing these clusters (membership Z score > 2). Of these, the figure shows the 59 genes that appear in a figure or table of Fain<sup>353</sup> these are genes whose functions are considered important for neuronal physiology.



**Supplementary Figure 9. Regional Association plots for Four Likely Causal SNPs Identified using CAVIARBF.**

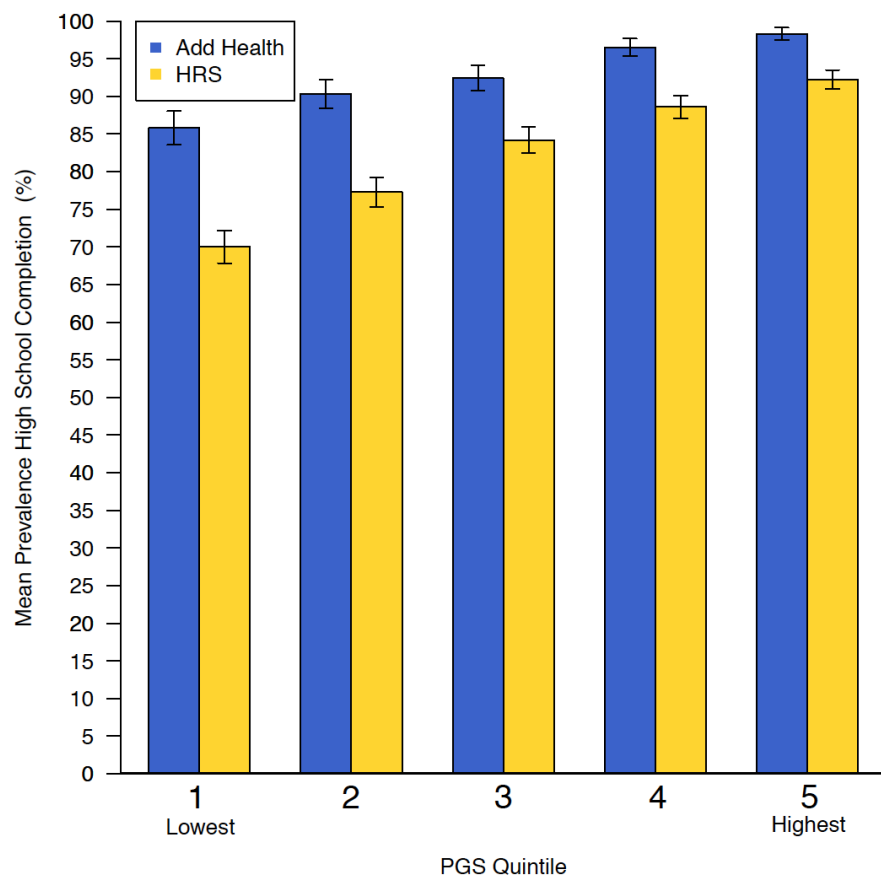
We show the top four SNPs, as assessed by posterior inclusion probability, that are both nonsynonymous and located in a DEPICT-prioritized gene.



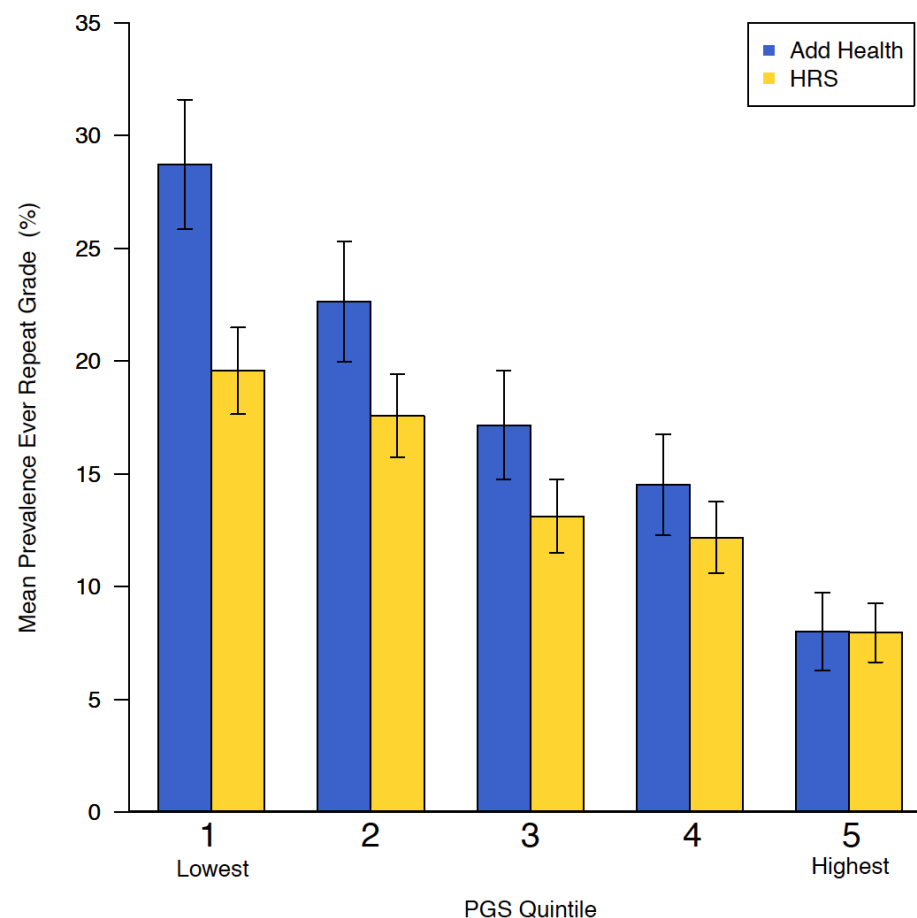
**Supplementary Figure 10. Predictive Power of Polygenic Score as a Function of Pruning at Different  $P$  Value Thresholds.**

Each PGS is based on a set of approximately independent SNPs identified using the clumping algorithm defined in **Supplementary Section 1.8**. For *HRS* ( $N = 8,609$ ) and *Add Health* ( $N = 4,775$ ) respectively, the number of SNPs included in the PGS is (with  $P$  value threshold in parentheses): 1,235 and 1,043 ( $5 \times 10^{-8}$ ); 5,152 and 4,432 ( $5 \times 10^{-5}$ ); 23,659 and 21,473 ( $5 \times 10^{-3}$ ); 262,908 and 244,603 (All SNPs); 1,104,681 and 1,169,298 (All SNPs, LDpred).

(A)

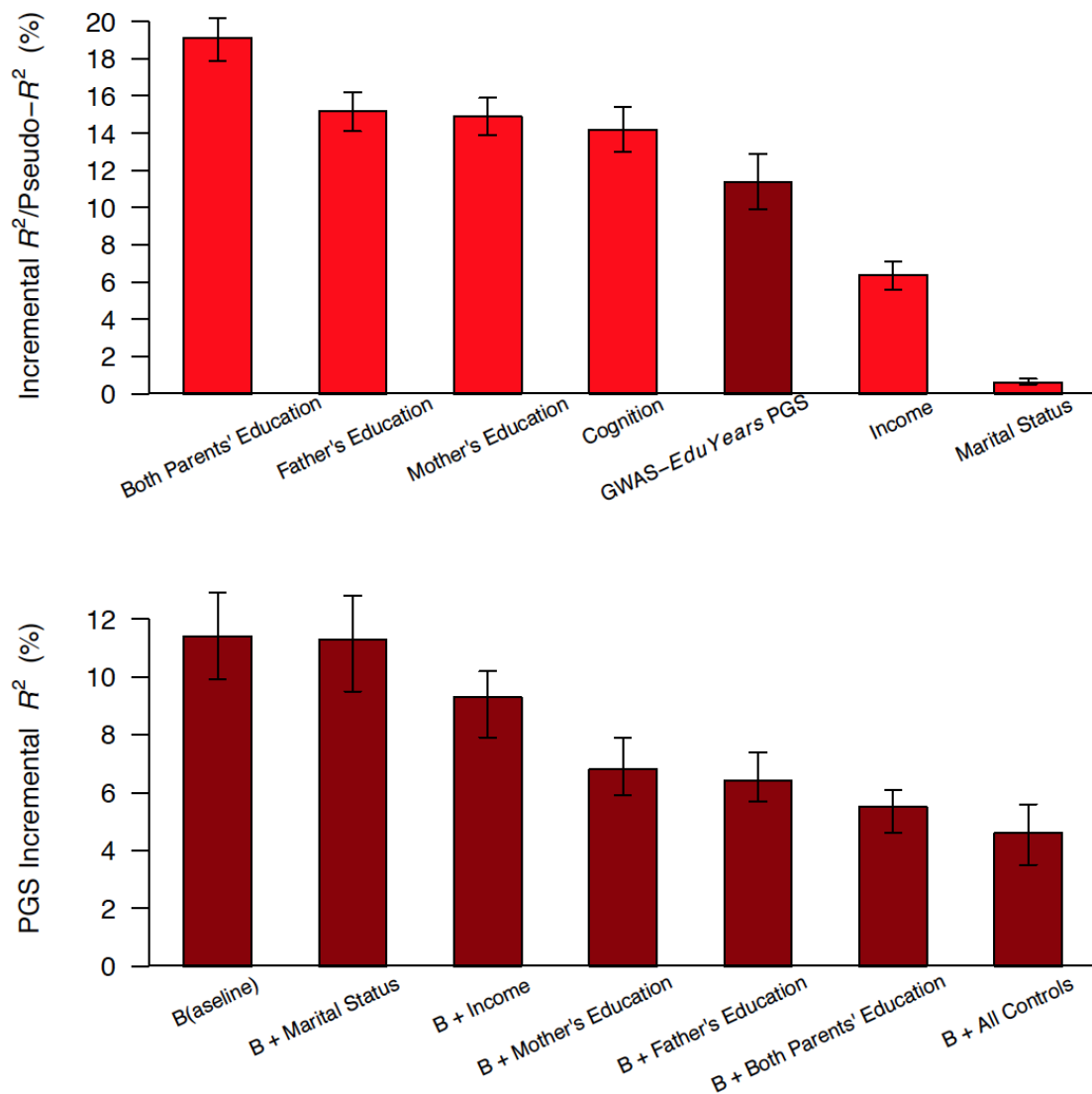


(B)



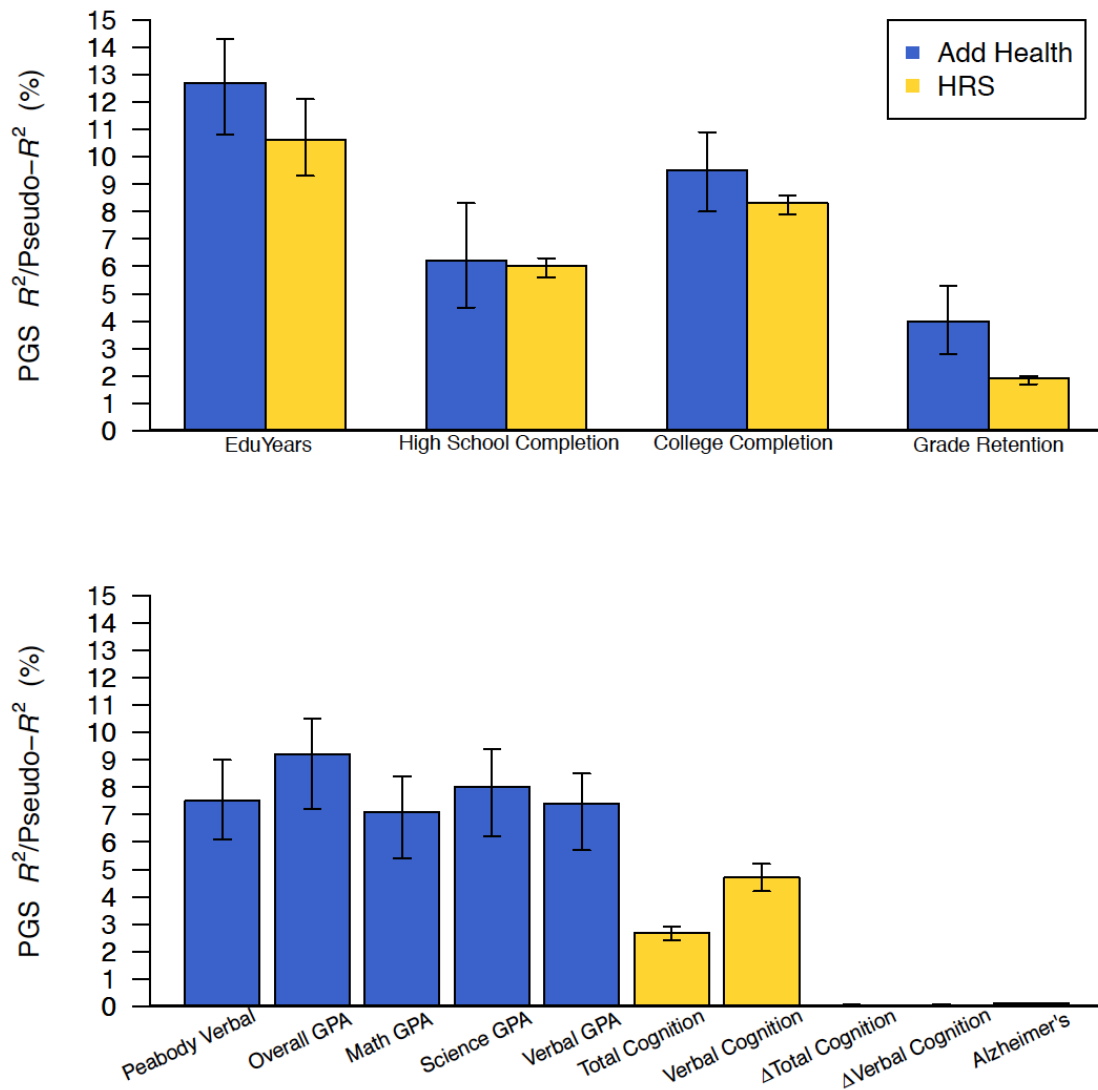
**Supplementary Figure 11. Mean Prevalence of Schooling Outcomes by *EduYears* PGS Quintile.**

Each quintile contains approximately 1,600 respondents in *HRS* and 900 respondents in *Add Health*. Total sample sizes for these two phenotypes for each of these prediction cohorts can be found in Supplementary Table 38. Quintile 1 contains the lowest PGS values; Quintile 5, the highest. Error bars show 95% confidence intervals. (A) High school completion, (B) Grade retention.



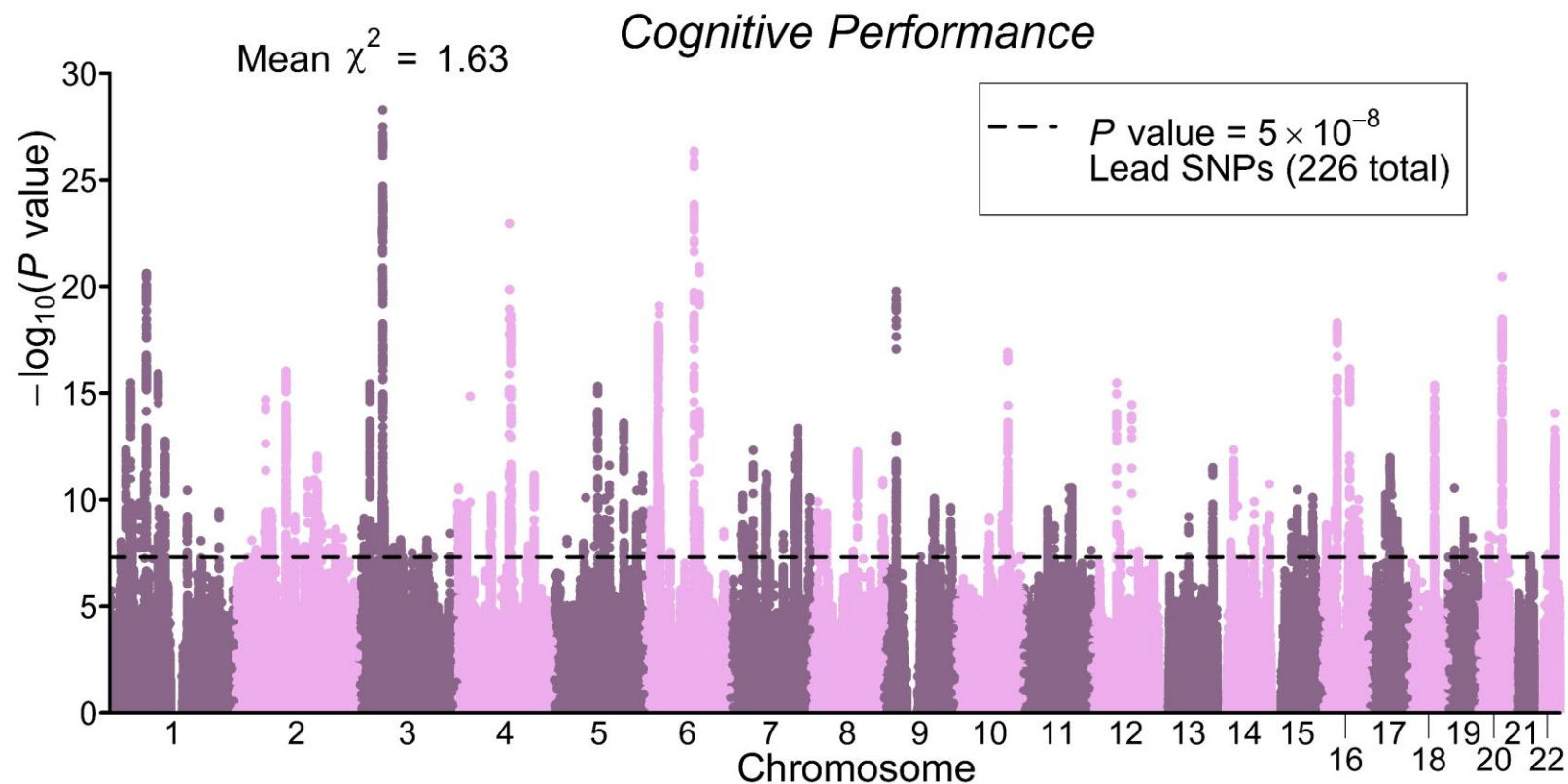
**Supplementary Figure 12. Predictive Power of GWAS-EduYears Polygenic Score Compared to Other Variables (top) and as Attenuated by Additional Controls (bottom).**

The outcome variable for all analyses is *EduYears*. In the top panel, incremental  $R^2$  values are calculated for each listed variable (or variables, in the case of both parents' education) and can be compared to the incremental  $R^2$  of the *EduYears* PGS, shown in dark red. All analyses in the top panel include the baseline control variables: sex, birth year, the interaction between sex and birth year, and the first ten principal components (PCs) of the genetic relatedness matrix. In the bottom panel, the left-most bar ("Baseline") is the same as the dark red bar from the top panel. Each of the other bars is the incremental  $R^2$  of the *EduYears* PGS, after controlling for the variables listed underneath the bar. Each bar corresponds to a sample-size weighted meta-analysis of the incremental  $R^2$  in both the *Add Health* and *HRS* cohorts, for a combined prediction sample size of 13,384.



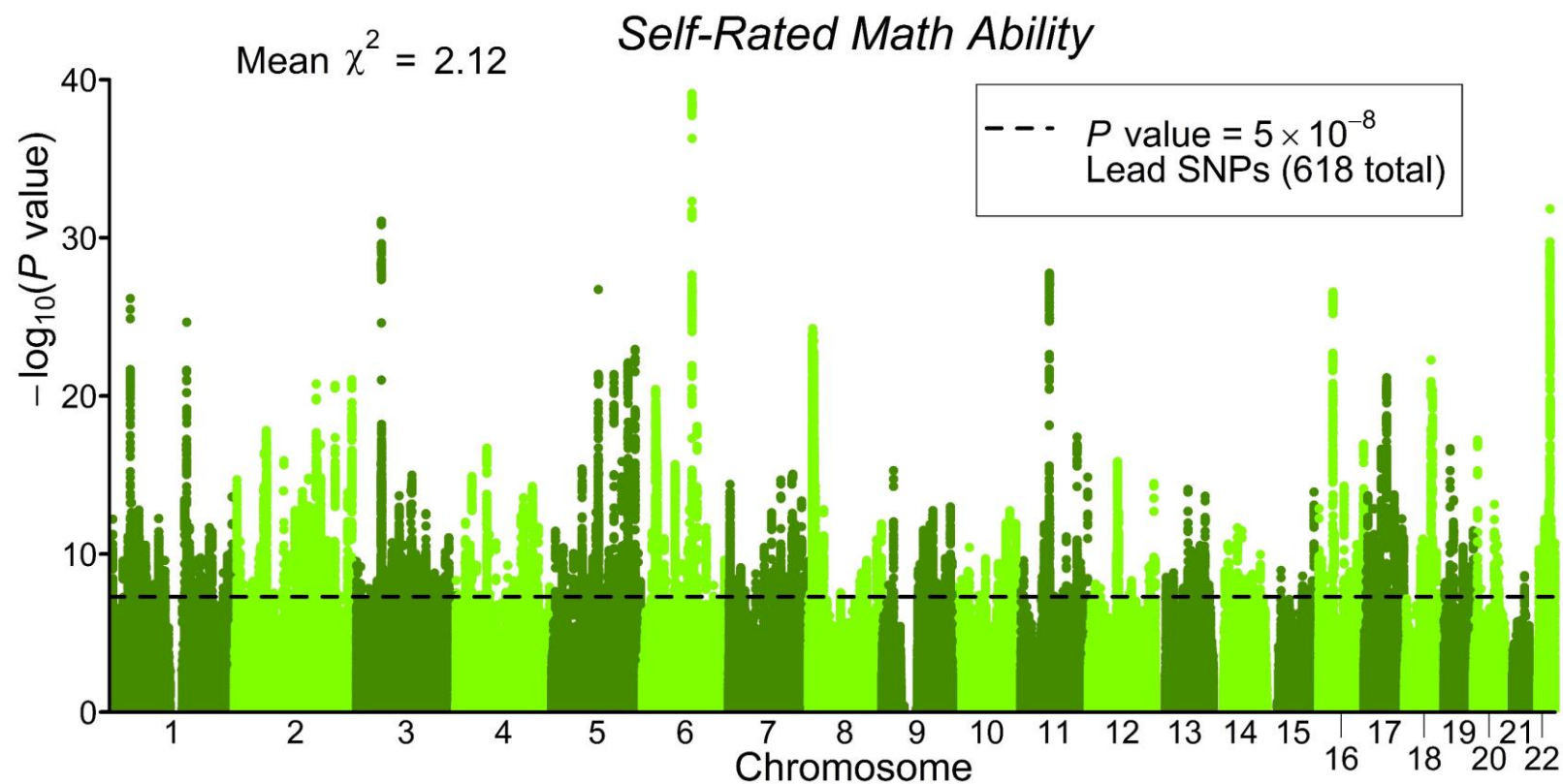
### Supplementary Figure 13. Polygenic Score Prediction in *Add Health* and *HRS*.

Predictive power of the polygenic score constructed from the current *EduYears* GWAS results in two independent prediction cohorts (*Add Health* and *HRS*). Panel A displays results for education phenotypes available in both datasets, and Panel B displays results for cognitive and academic achievement phenotypes available in either *Add Health* or *HRS*. Error bars show bootstrapped 95% confidence intervals with 1,000 iterations each. The number of individuals in the prediction sample for each regression represented above can be found in **Supplementary Table 38**.



**Supplementary Figure 14. Manhattan Plot for *Cognitive Performance* ( $N = 257,841$ ).**

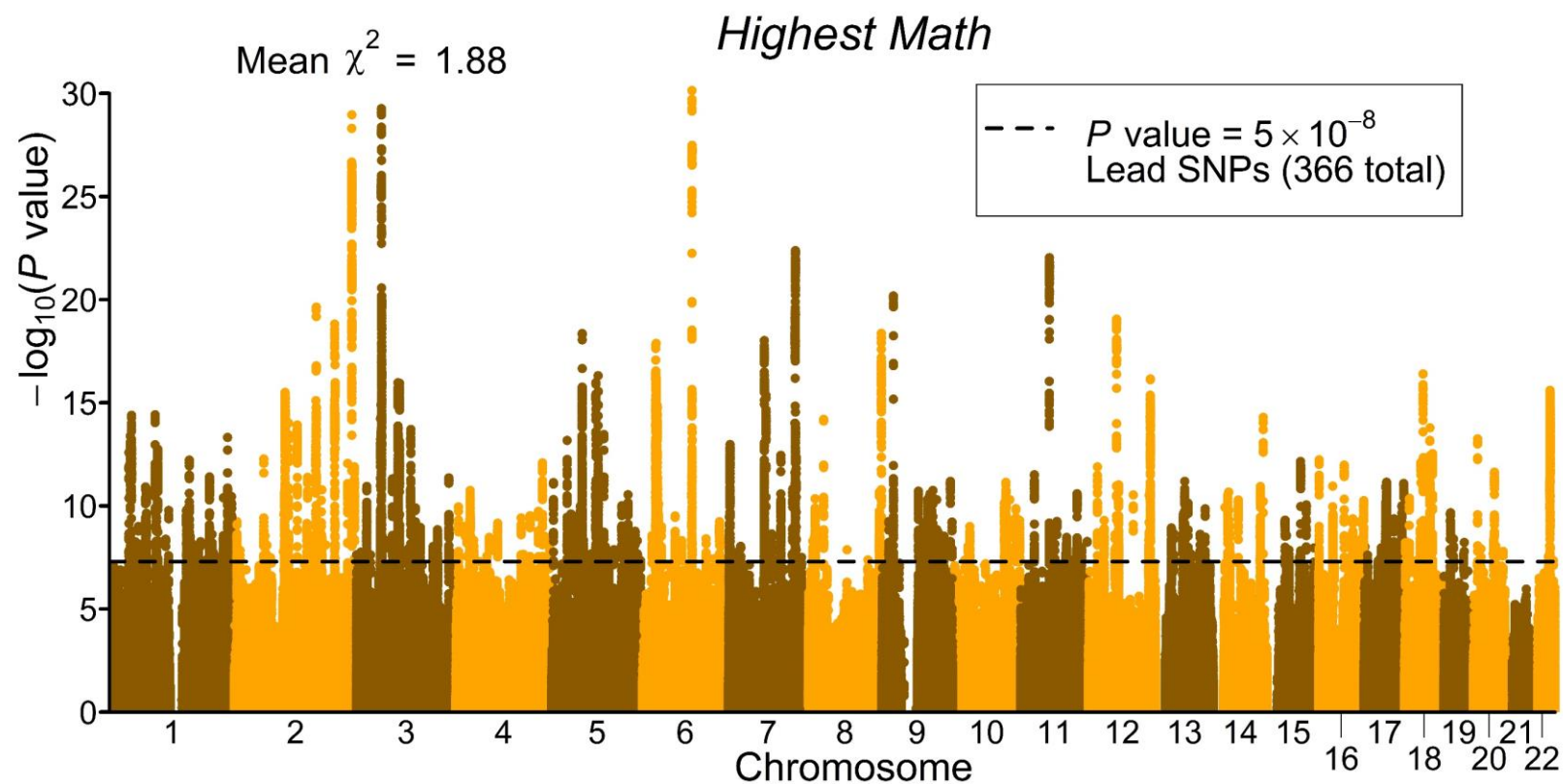
SNPs are plotted on the  $x$ -axis according to their position on each chromosome, and significance of association with the phenotype is on the  $y$ -axis [shown as  $-\log_{10}(P \text{ value})$ ]. The solid line indicates the threshold for genome-wide significance ( $P = 5 \times 10^{-8}$ ). All  $P$  values are derived from test statistics inflated by the estimated intercept from an LD Score regression.



**Supplementary Figure 15. Manhattan Plot for *Self-Rated Math Ability* ( $N = 564,698$ ).**

SNPs are plotted on the  $x$ -axis according to their position on each chromosome, and significance of association with the phenotype is on the  $y$ -axis [shown as  $-\log_{10}(P \text{ value})$ ]. The solid line indicates the threshold for genome-wide significance ( $P = 5 \times 10^{-8}$ ). All  $P$  values are derived from test statistics inflated by the estimated intercept from an LD Score regression.

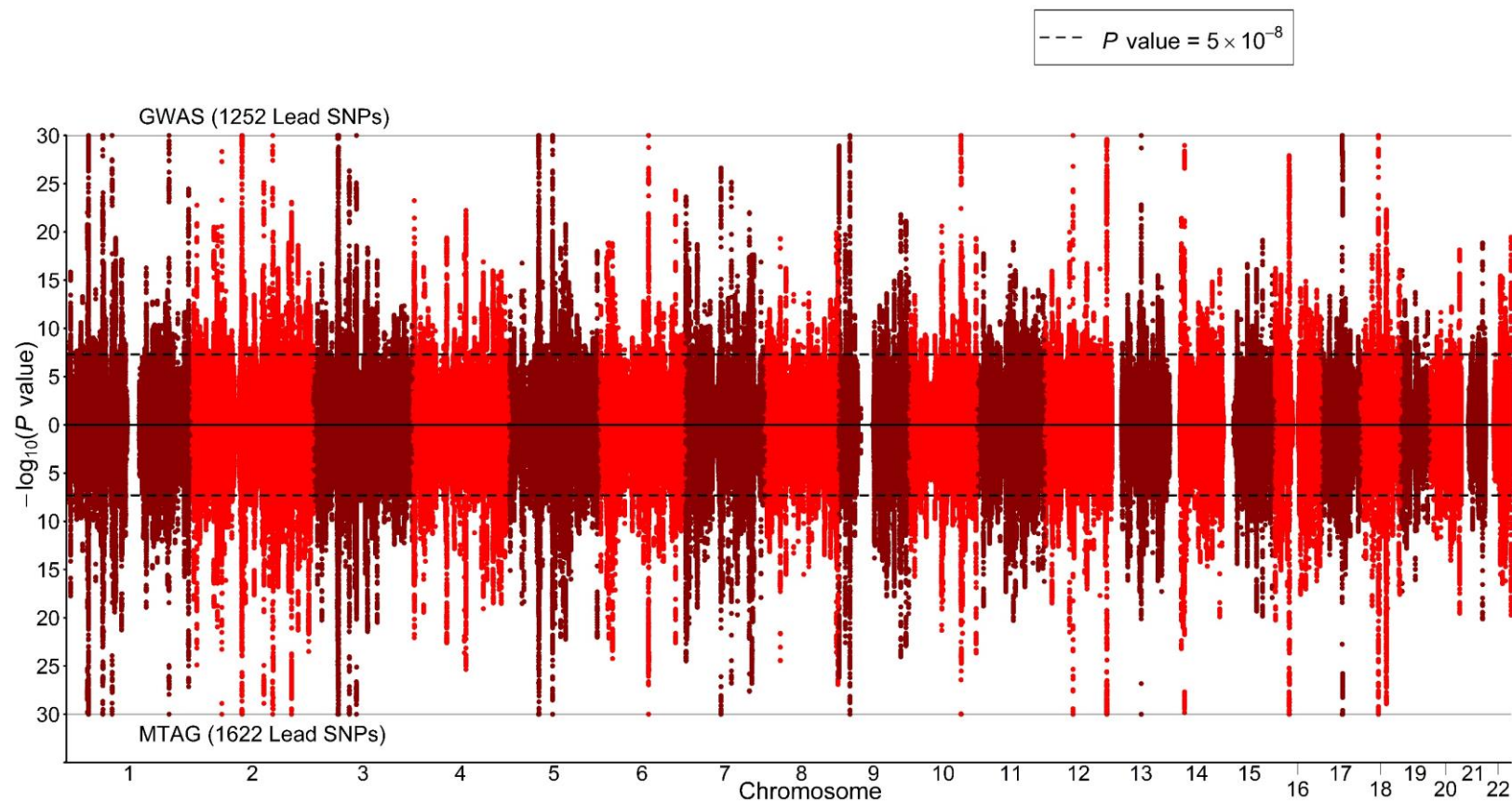




**Supplementary Figure 16. Manhattan Plot for *Highest Math* ( $N = 430,445$ ).**

SNPs are plotted on the  $x$ -axis according to their position on each chromosome, and significance of association with the phenotype is on the  $y$ -axis [as  $-\log_{10}(P \text{ value})$ ]. The solid line indicates the threshold for genome-wide significance ( $P = 5 \times 10^{-8}$ ). All  $P$  values are derived from test statistics inflated by the estimated intercept from an LD Score regression.

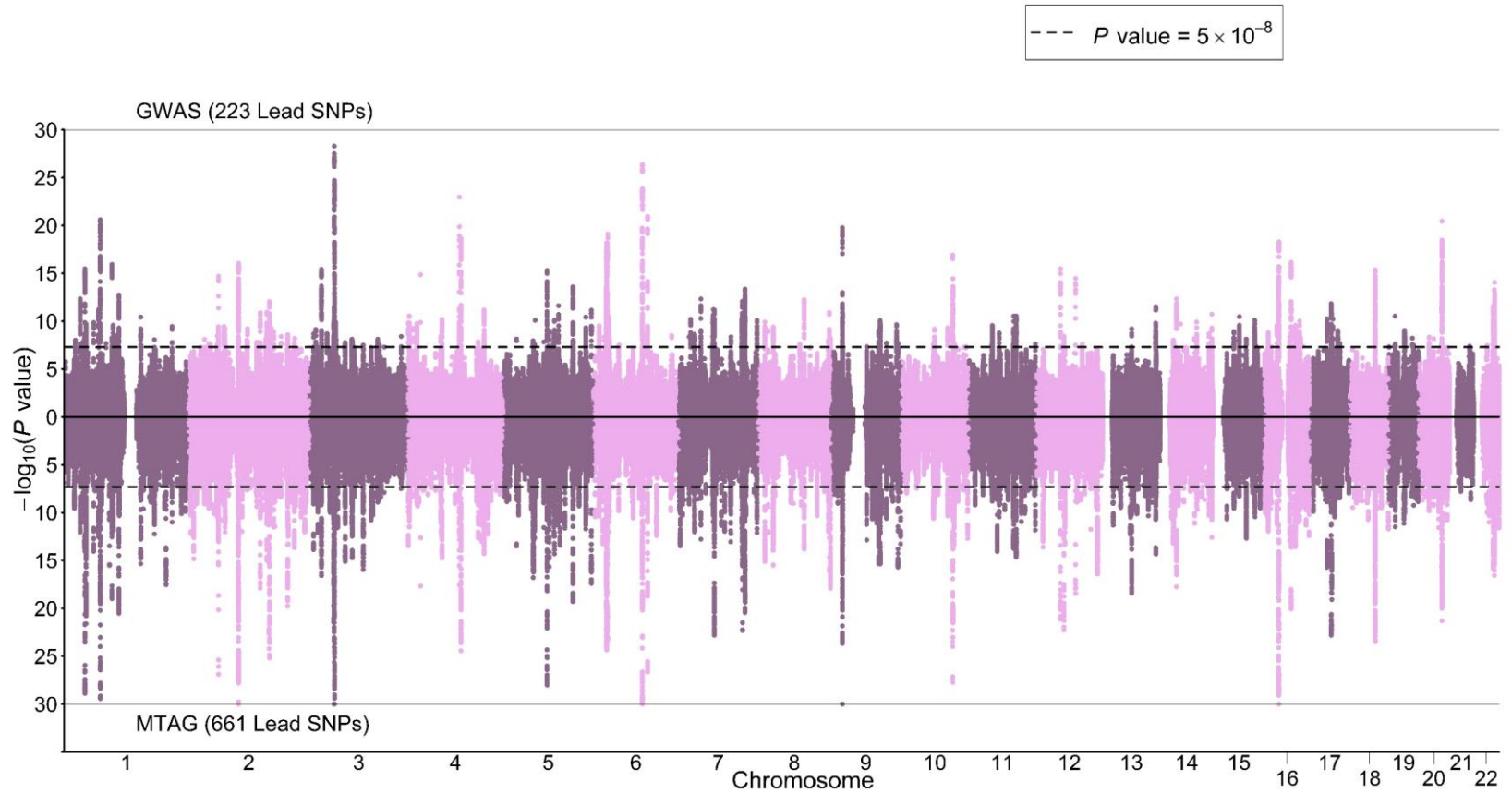
### EduYears (GWAS vs. MTAG)



**Supplementary Figure 17. Inverted Manhattan Plot of GWAS and MTAG results for *EduYears*.**

To facilitate comparisons, the GWAS and MTAG results are shown for the set of ~7M SNPs that passed MTAG filters. The average  $\chi^2$ -statistic reported is calculated based on adjusted test statistics. The  $-\log_{10}(P \text{ value})$  is truncated at  $\pm 30$ .

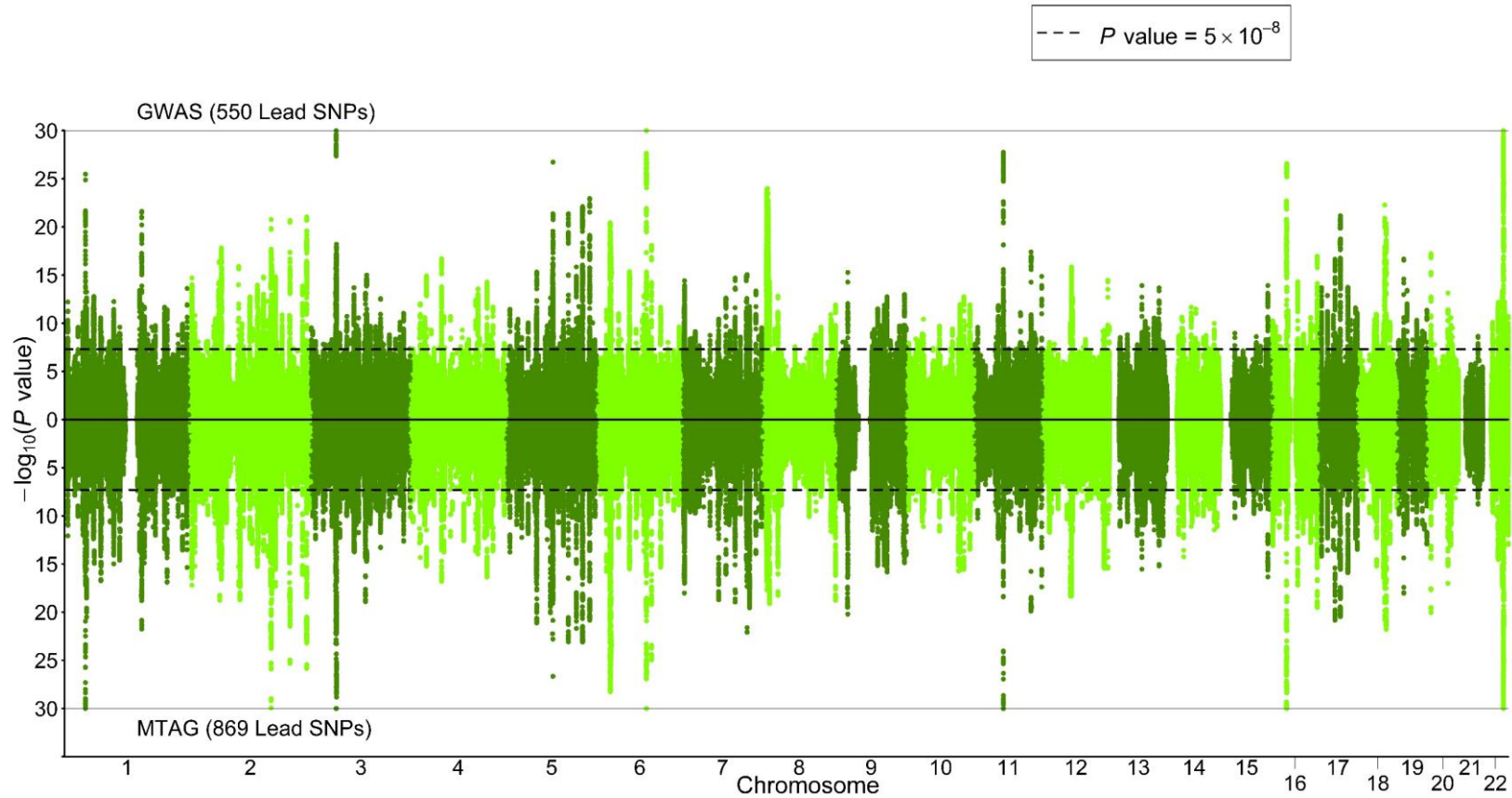
### Cognitive Performance (GWAS vs. MTAG)



**Supplementary Figure 18. Inverted Manhattan Plot of GWAS and MTAG Results for *Cognitive Performance*.**

To facilitate comparisons, the GWAS and MTAG results are shown for the set of ~7M SNPs that passed MTAG filters. The average  $\chi^2$ -statistic reported is calculated based on adjusted test statistics. The  $-\log_{10}(P \text{ value})$  is truncated at  $\pm 30$ .

### Self-Rated Math Ability (GWAS vs. MTAG)

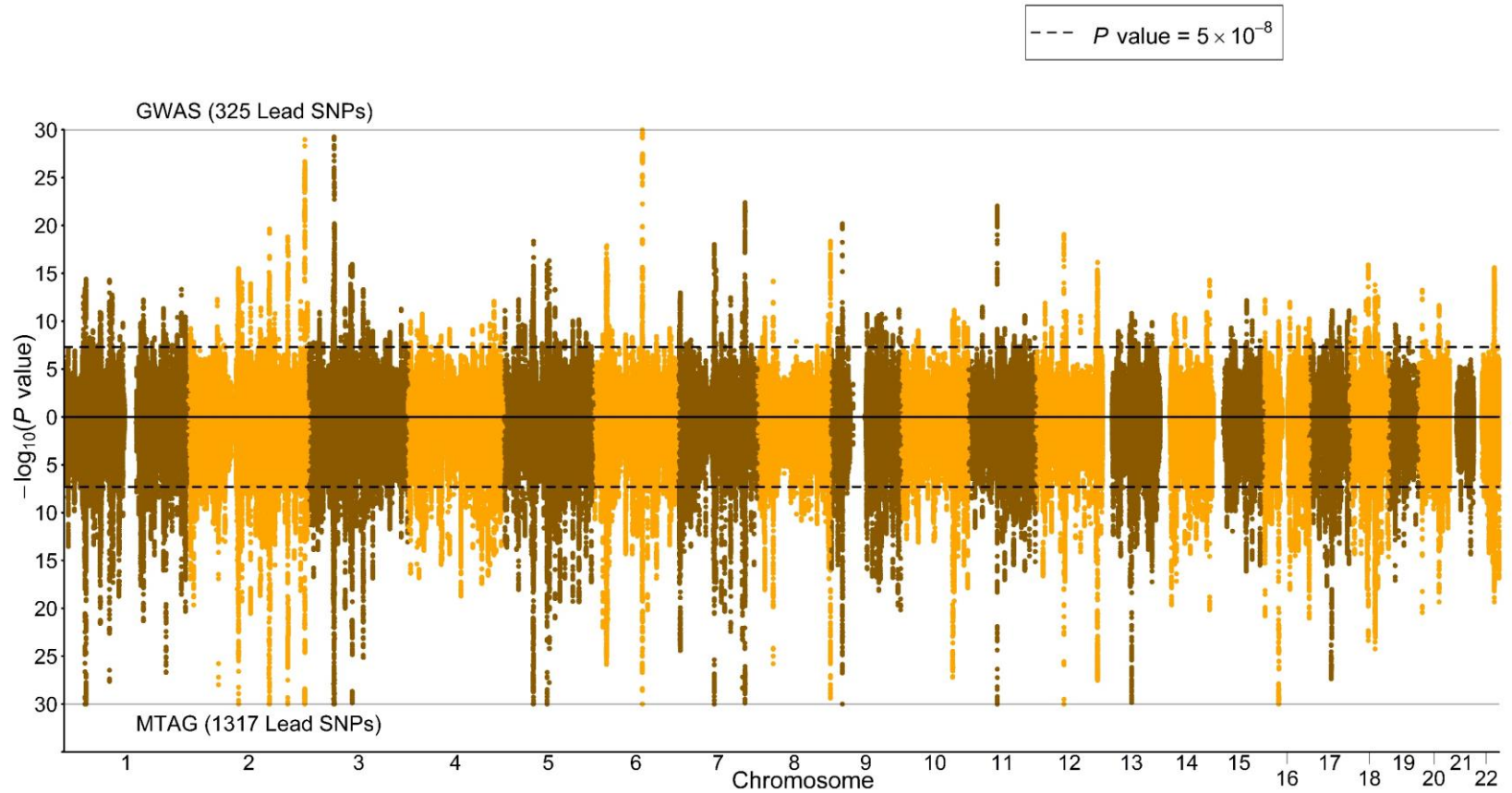


**Supplementary Figure 19. Inverted Manhattan Plot of GWAS and MTAG Results for *Math Ability*.**

To facilitate comparisons, the GWAS and MTAG results are shown for the set of ~7M SNPs that passed MTAG filters. The average  $\chi^2$ -statistic reported is calculated based on adjusted test statistics. The  $-\log_{10}(P \text{ value})$  is truncated at  $\pm 30$ .

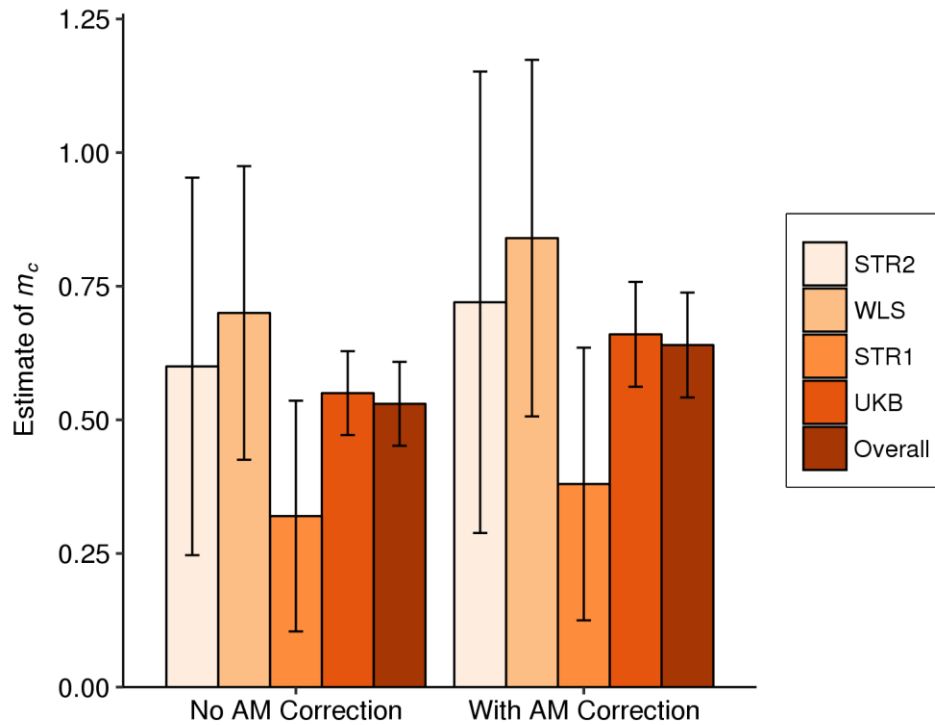


### Highest Math (GWAS vs. MTAG)



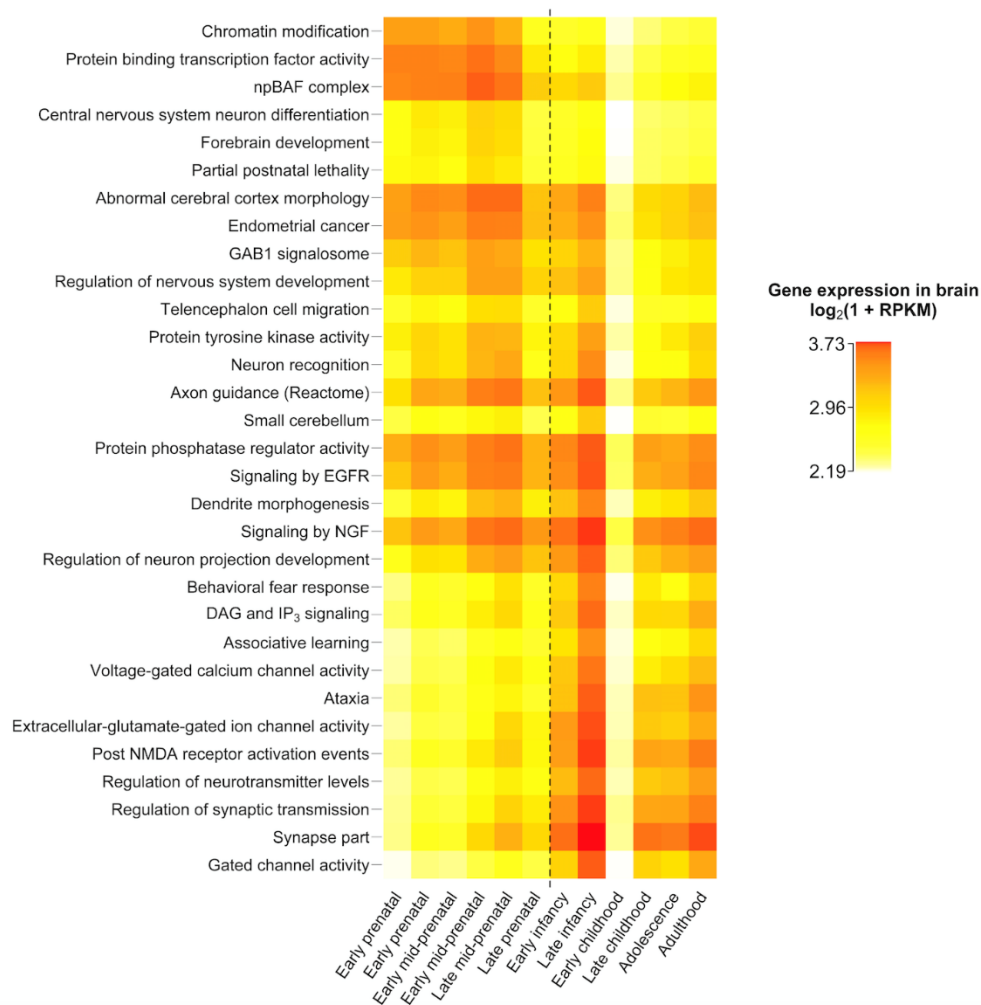
**Supplementary Figure 20. Inverted Manhattan Plot of GWAS and MTAG Results for *Highest Math*.**

To facilitate comparisons, the GWAS and MTAG results are shown for the set of ~7M SNPs that passed MTAG filters. The average  $\chi^2$ -statistic reported is calculated based on adjusted test statistics. The  $-\log_{10}(P \text{ value})$  is truncated at  $\pm 30$ .



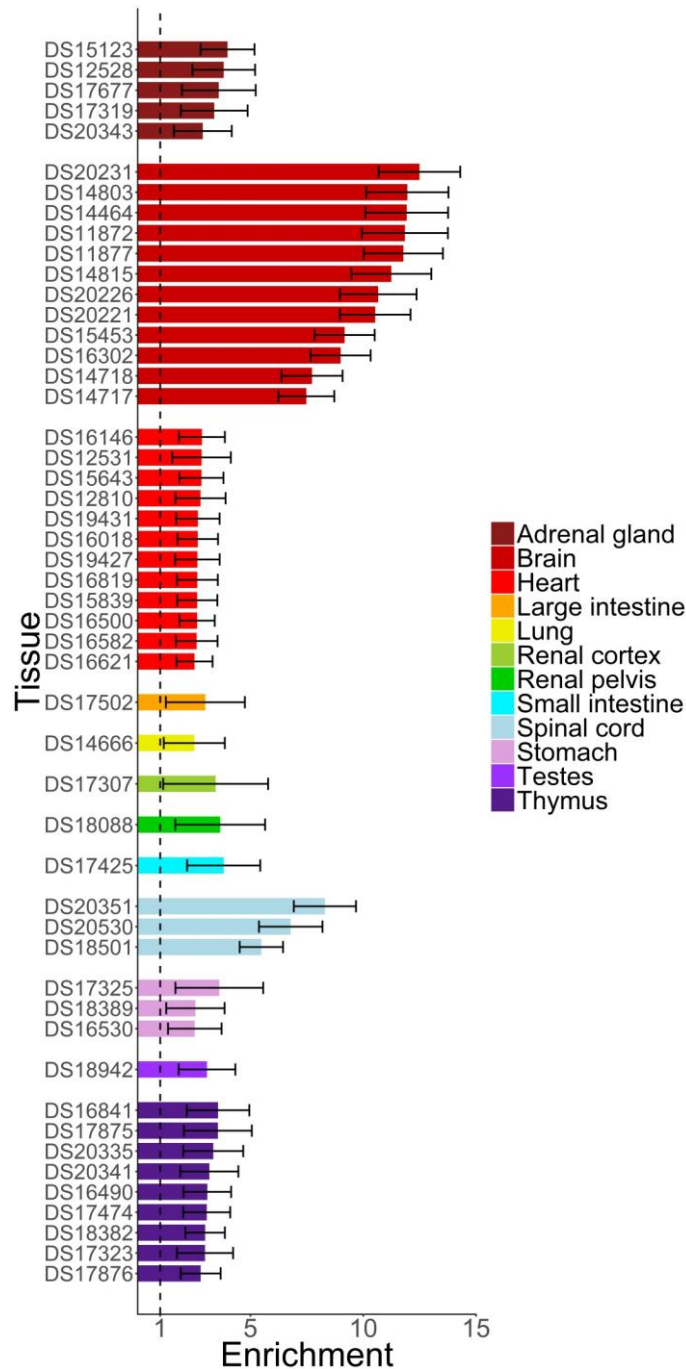
**Supplementary Figure 21. Summary Overview of  $m_c$  Estimates in Sibling Cohorts.**

Each estimate of  $m_c$  was calculated by comparing *EduYears* summary statistics from a between-family GWAS with those from a within-family analysis. The between-family estimates were calculated in an identical way to the main GWAS results but excluding siblings and their relatives. Cohorts are ordered by size of the sibling sample (smallest to largest), with “Overall” using within-family summary statistics from an inverse-variance-weighted meta-analysis of the four cohorts. The sample sizes for the discovery and within-family cohorts for each estimate represented above are found in **Supplementary Table 21**. Error bars show 95% confidence intervals calculated using the block jackknife with 1000 iterations.



### Supplementary Figure 22. Brain-Specific Expression of Significantly Enriched Gene Sets across Development.

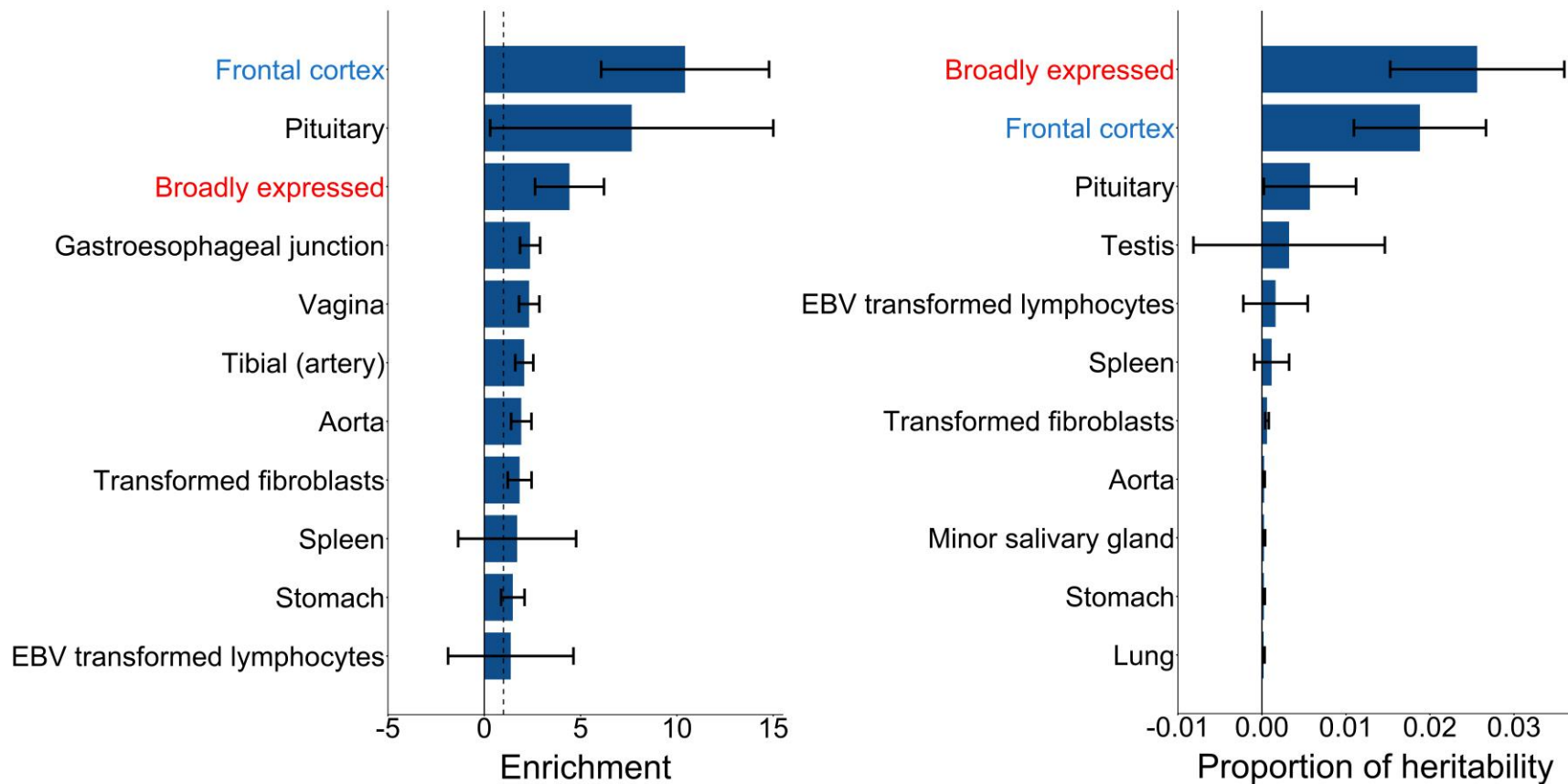
Each row is a non-PPI gene set prioritized by DEPICT ( $\text{FDR} < 0.05$ ) and chosen as the exemplary member of its cluster by the Affinity Propagation algorithm, based on the results of the autosomal GWAS of *EduYears*. The colors represent the expression of the gene set in the BrainSpan Developmental Transcriptome as a function of developmental stage. Expression of the gene set is calculated as the weighted mean of the expression of all DEPICT-prioritized genes ( $\text{FDR} < 0.05$ ), with each gene's weight derived from its membership score. The gene sets are ordered by the difference between the mean prenatal and postnatal expression. We show only those sets with a heritability enrichment greater than 1.25 according to stratified LD Score regression; complete results can be found in **Supplementary Table 8**. (Note that the **early childhood** stage appears to be an outlier, with uniformly low expression across gene sets.)



**Supplementary Figure 23. DNase I Hypersensitivity in Fetal Tissues/Cell Types as a Predictor of SNP Effects on *EduYears*.**

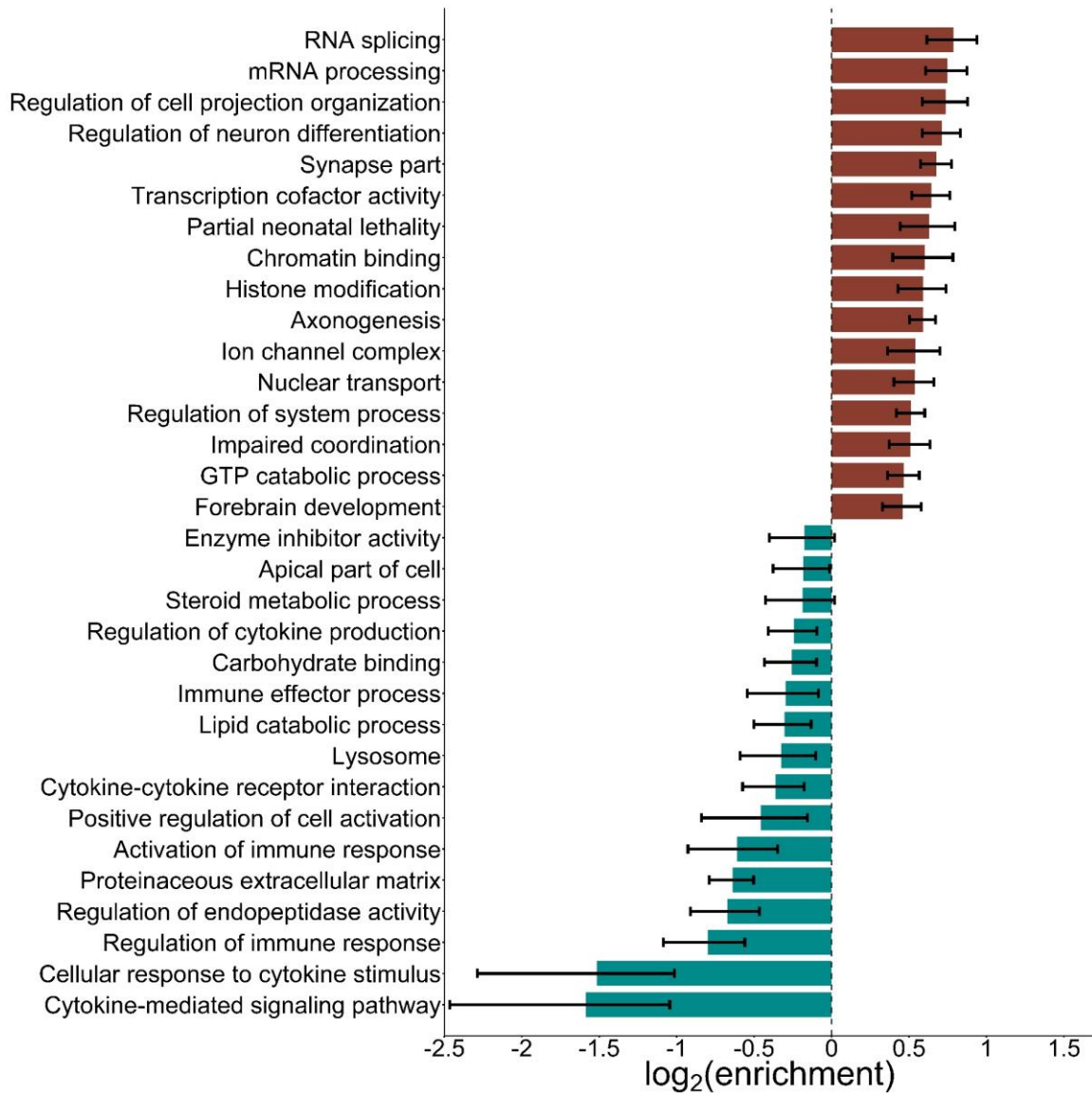
We applied stratified LD Score regression<sup>24</sup>, as in the analysis whose results are displayed in **Supplementary Figure 25**, except this time using SNP-level annotations from Pickrell<sup>60</sup>. We display the top 50 estimated enrichment factors and associated 95% confidence intervals of annotations referring to DNase I hypersensitivity in a fetal tissue/cell type. Complete results can be found in **Supplementary Table 32**.





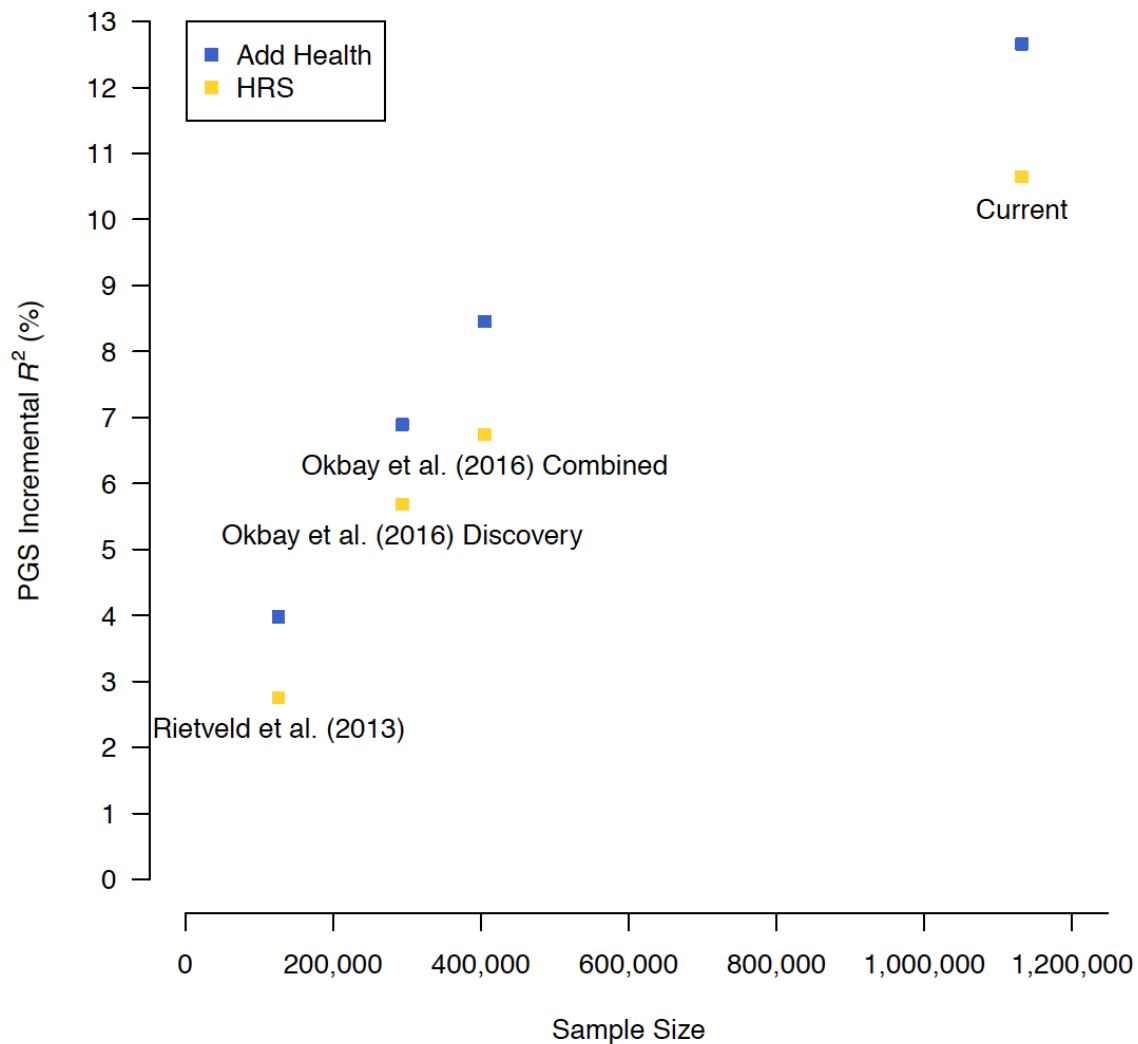
**Supplementary Figure 24. Heritability Enrichment of Genes That Are Broadly or Specifically Expressed.**

Genes with brain-specific expression show the strongest enrichment of *EduYears* heritability (left), but genes broadly expressed across many tissues contribute more to total heritability due to their greater number (right). We used stratified LD Score regression<sup>24</sup> to partition the heritable variation. Error bars represent 95% confidence intervals. For comparability with Boyle et al.<sup>332</sup> we display the top 11 tissues in each panel. Complete results can be found in **Supplementary Table 34**.



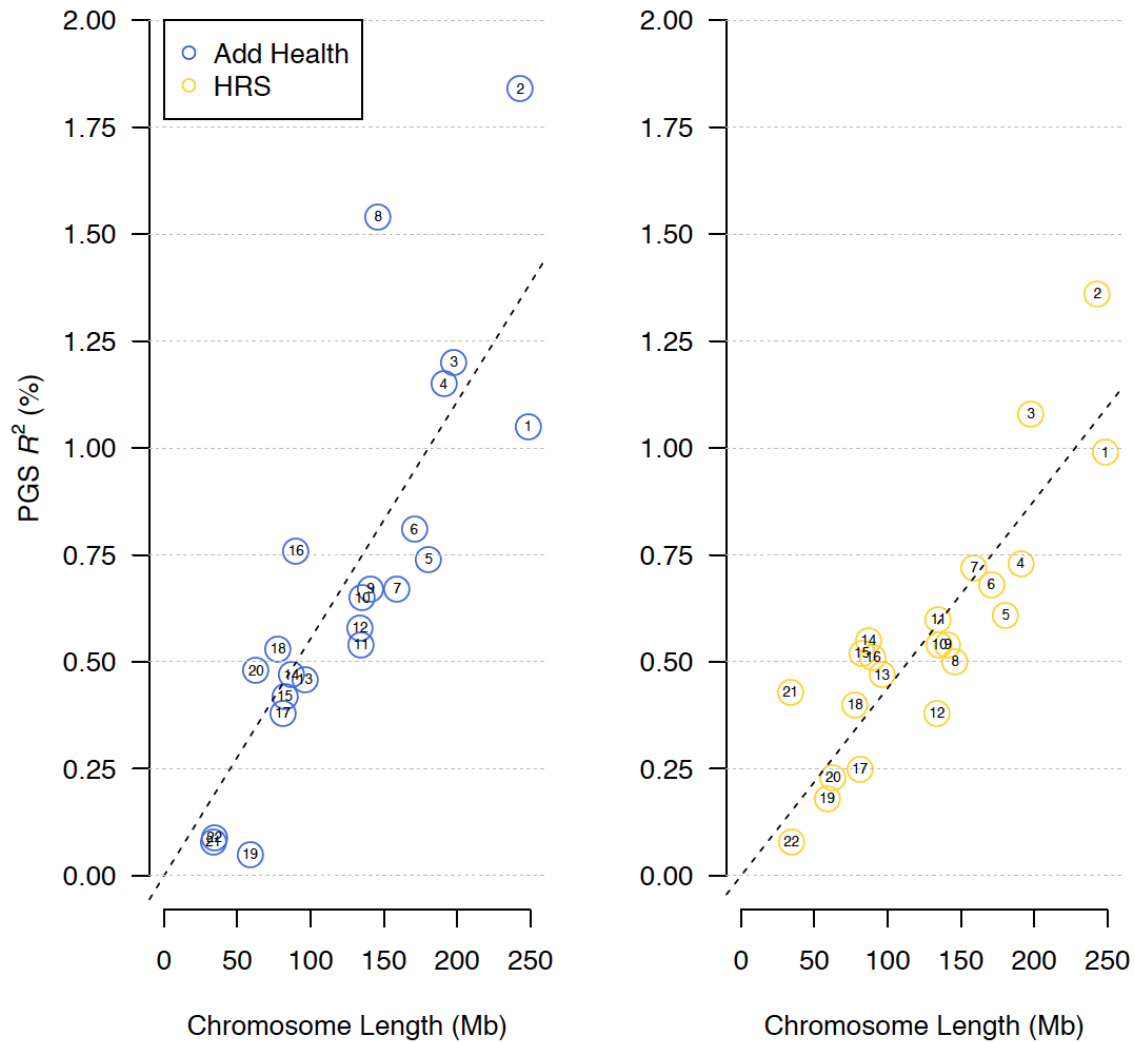
**Supplementary Figure 25. Binary Gene Sets with Strongest and Weakest Heritability Enrichment (15 of Each).**

The gene sets are the original, binary (not reconstituted) gene sets. The enrichment factor for a given gene set is the ratio of two quantities: (1) the numerator is the fraction of the *EduYears* heritability explained by SNPs mapping to genes that are members of the set, and (2) the denominator is the fraction of SNPs that map to those genes. We selected those (original, binary) non-PPI gene sets with reconstituted versions found by DEPICT to be significantly enriched ( $FDR < 0.05$ ) or impoverished ( $P$  value = 1) and with at least 200 original members. We used stratified LD Score regression<sup>24</sup> to partition the heritable variation. Error bars represent  $\pm 1$  standard error. We show the 15 enriched gene sets with the largest enrichment factors (red), subject to the constraint that each lower-ranking result comes from a different cluster in **Supplementary Table 8** than all higher-ranking results. We also show the 15 impoverished sets with the smallest factors (green), subject to an analogous constraint. Complete results can be found in **Supplementary Table 36**.



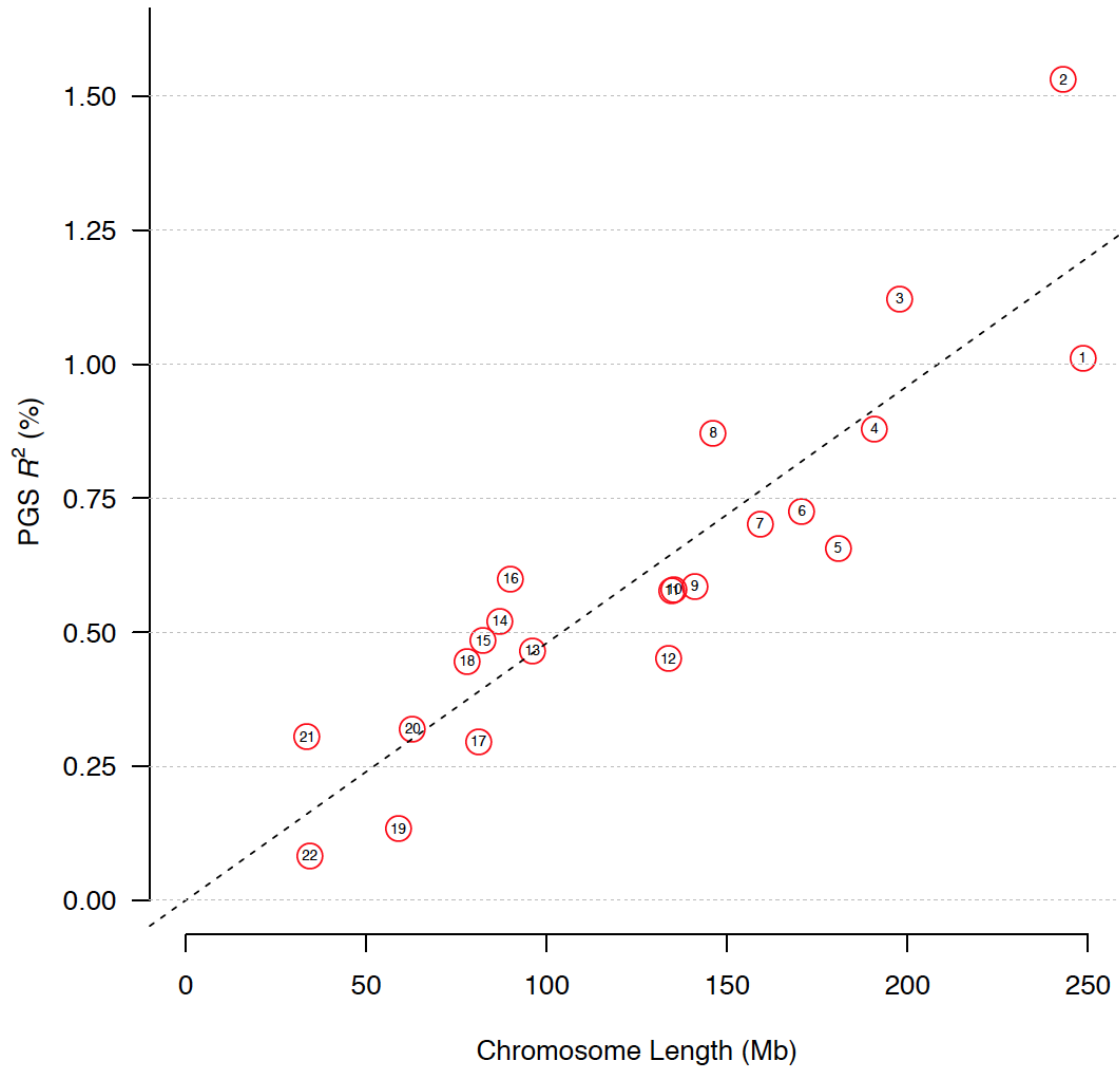
**Supplementary Figure 26. Predictive Power of Polygenic Score as a Function of the Size of the *EduYears* GWAS Discovery Sample.**

All PGSs constructed using the software LDpred<sup>354</sup> assuming a normal prior for SNP effect sizes. Prediction samples are European-ancestry subjects in the National Longitudinal Study of Adolescent to Adult Health (*Add Health*,  $N = 4,775$ ) and the Health and Retirement Study (*HRS*,  $N = 8,609$ ). Incremental  $R^2$  is the difference between the  $R^2$  from a regression of *EduYears* on the PGS and the controls (sex, age, their interaction, and 10 PCs) and the  $R^2$  from a regression of *EduYears* on just the controls. All PGSs are constructed from *EduYears* GWAS results that exclude *Add Health* and *HRS*.



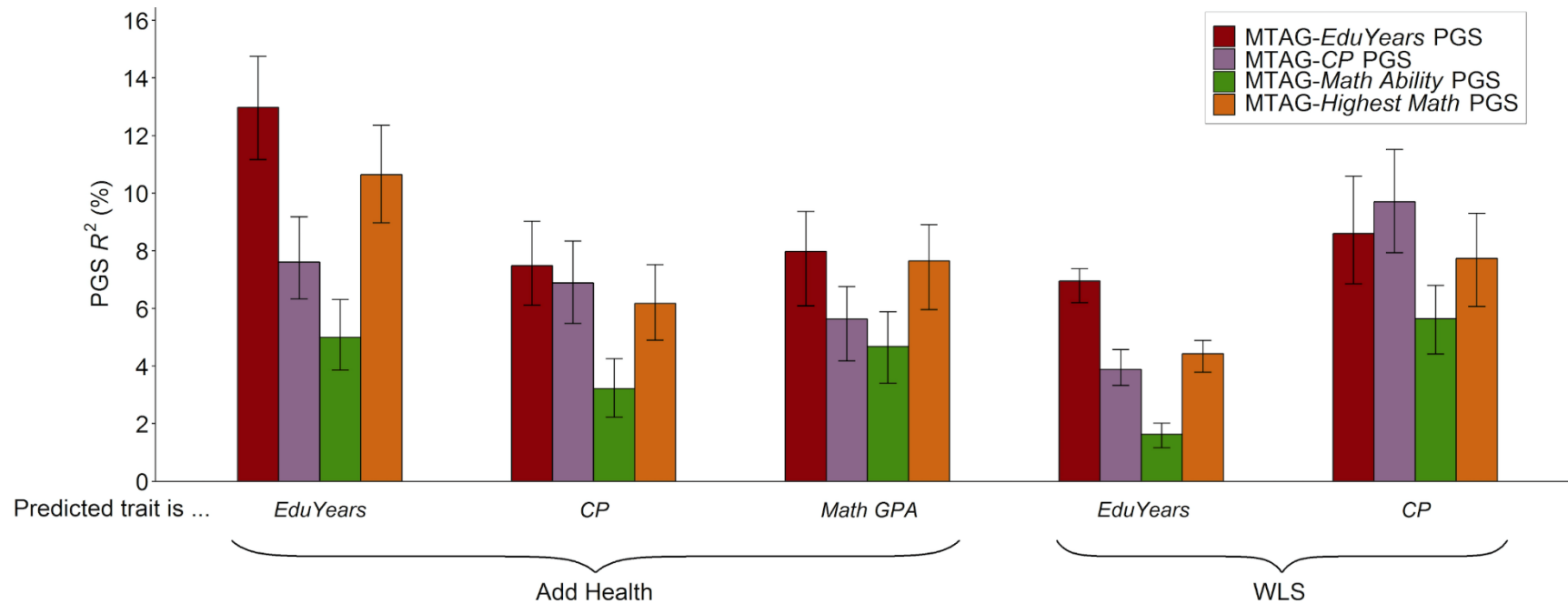
**Supplementary Figure 27. Predictive Power of Chromosome-Specific *EduYears* Polygenic Scores in *Add Health* and *HRS*.**

The dashed line depicts the best fit from a regression of the chromosome-specific polygenic score's incremental  $R^2$  on chromosome length, with the intercept constrained to zero. The prediction sample sizes for the *Add Health* and *HRS* cohorts are 4,775 and 8,609, respectively.



**Supplementary Figure 28. Predictive Power of Chromosome-Specific *EduYears* Polygenic Scores in Sample-Size Weighted Meta-Analysis of *Add Health* and *HRS*.**

The dashed line depicts the best fit from a regression of the chromosome-specific polygenic score's incremental  $R^2$  on chromosome length, with the intercept constrained to zero. The combined prediction sample size of the two cohorts is 13,384.



**Supplementary Figure 29. Comparison MTAG PGSs Based on Trait-Specific MTAG Association Statistics and MTAG Association Statistics for Other Traits.**

In both prediction cohorts, the MTAG-*EduYears* score is the best predictor of *EduYears*, and MTAG-*CP* is the best predictor of cognitive performance. In *Add Health*, cognitive performance is the respondent's age-standardized score on a test of verbal cognition. In *WLS*, cognitive performance is defined as the respondent's score on a Henmon-Nelson test of mental ability<sup>352</sup>. Math GPA is a proxy for *Math Ability* and *Highest Math*. Error bars show bootstrapped 95% confidence intervals with 1,000 iterations each. The number of individuals in the prediction sample for each regression represented above can be found in **Supplementary Table 43**.